

Chapter 4

Formal and Pragmatic Aspects of Conceptual Covers

In front of you lie two cards. One is the Ace of Spades, the other is the Ace of Hearts. Their faces are turned over. You do not know which one is which. There are two different ways of identifying the two cards in our domain, namely by their position on the table (the card on the left, the card on the right) and by their suit (the Ace of Spades, the Ace of Hearts). These identification methods are typical examples of what I have called conceptual covers in the previous chapters. In this chapter, I investigate the notion of a conceptual cover in more detail. In section 4.1, I restate the definition and discuss some general properties. I then investigate which view of trans-world identification is formalized by the notion of a conceptual cover and compare it with other views that have been proposed in the literature (section 4.2). Finally I will attempt a first analysis of the pragmatic procedures of conceptual cover selection (section 4.3).

4.1 Conceptual Covers

Given a set of worlds W and a set of individuals D , an *individual concept* is a total function from W to D . I call the *instantiation* of c in w the value $c(w) = d$ that a concept c assigns to a world w . A conceptual cover is a set of individual concepts which satisfies the following condition: in each world, each individual constitutes the instantiation of one and only one concept.

4.1.1. DEFINITION. Let W be a set of possible worlds and D a set of individuals. A *Conceptual Cover* CC over (W, D) is a set of individual concepts such that:

- (i) $\forall w \in W : \forall d \in D : \exists c \in CC : c(w) = d$;
- (ii) $\forall w \in W : \forall c, c' \in CC : c(w) = c'(w) \Rightarrow c = c'$.

The existential condition says that in a cover, each individual is identified by means of at least one concept in each world, that is, in each world, each individual should be accessible in some way. The uniqueness condition says that in no world, an individual is the instantiation of more than one concept, that is, in no world an individual is counted twice. This means that in a cover, either c_1 and c_2 are identical (their values coincide everywhere) or they are separated (their values never coincide). In a conceptual cover, each individual in the universe of discourse is identified in a determinate way, and different conceptual covers constitute different ways of conceiving of one and the same domain.

In the following two sections we prove two results, which have already been discussed in a number of applications in the previous chapters. The first proposition says that there is a one-to-one correspondence between a conceptual cover and the domain of individuals; the second proposition says that the number of covers definable over a set of worlds and a set of individuals is relative to the cardinality of these two sets. Technically these two results are not so impressive, but their intuitive consequences are interesting.

one to one

There is a one-to-one correspondence between a conceptual cover and the domain of individuals.

4.1.2. PROPOSITION. Let W be a set of worlds and D a set of individuals. Let CC be a conceptual cover based on (W, D) . Then $|CC| = |D|$.

proof: Let f be a function from CC to D such that for some world w , the following holds: $f(c) = c(w)$. It is enough to show that f is a bijection. That f is injective and surjective follow directly from the uniqueness and the existential condition respectively:

- (i) f is injective. $\forall c_1, c_2 \in CC : f(c_1) = f(c_2) \Rightarrow c_1(w) = c_2(w)$ (by construction) $\Rightarrow c_1 = c_2$ (by uniqueness condition).
- (ii) f is surjective. $\forall d \in D : \exists c \in CC : c(w) = d$ (by existence condition) $\Rightarrow \forall d \in D : \exists c \in CC : f(c) = d$ (by construction). \square

In a conceptual cover, each individual is identified by one and only one concept. Different conceptual covers constitute different ways of identifying individuals of the same sort. Irrespective of which perspective you assume, the number of individuals does not change.

A first consequence of proposition 4.1.2 is that in defining our models, we can drop the one domain assumption, as long as the domains associated with the individual worlds have the same cardinality.

Given this one to one correspondence between the universe of a model M and any conceptual cover based on M , any discussion about their relative priority

becomes vacuous. In chapter 2, section 2.4.4, we have proved the related result that given a model M and a conceptual cover CC based on M , we can always find another model M' which satisfies the same sentences of M in which the identification method formalized by CC is the rigid one. The contrast between rigid and non-rigid conceptual covers can be used to represent the contrast between demonstrative versus descriptive identification. In this perspective, the one-to-one result can be interpreted as saying that our logic remains neutral with respect to the debate between empiricists and rationalists about the relative priority of these two identification methods. In the construction of the objects of our experience, the direct ostensive moment can be prior to the descriptive conceptual moment or vice versa, or the two moments can be taken to presuppose each other. We are free to embrace any of these philosophical positions. Logically speaking, it does not matter.

relativity

The number of conceptual covers definable over a set of worlds and a set of individuals crucially depends on the cardinality of the two sets. There are $(|D|!)^{|W|-1}$ conceptual covers over (W, D) .¹

4.1.3. PROPOSITION. Let $C(W, D) = \{CC \mid CC \text{ is a cover based on } (W, D)\}$. Then $|C(W, D)| = (|D|!)^{|W|-1}$.

proof: the proof is by induction on the number of worlds $|W| = n$.

($n = 1$) Let $|W| = 1$. Suppose $W = \{w\}$. Then there is a unique cover based on (W, D) , namely $CC = \{\langle w, d \rangle \mid d \in D\}$. Thus $|C(W, D)| = 1 = (|D|!)^0 = (|D|!)^{|W|-1}$.

($n \Rightarrow n + 1$) Let $|W| = n + 1$. We have to prove that $|C(W, D)| = (|D|!)^n$. Suppose $W = \{w_1, \dots, w_n, w_{n+1}\}$. Consider $W^* = \{w_1, \dots, w_n\}$ i.e. $W^* = W - \{w_{n+1}\}$. By induction hypothesis $|C(W^*, D)| = (|D|!)^{n-1}$. We have to check how many more conceptual covers we can have if we add w_{n+1} to W^* . Let $|D| = m$ and $CC^* = \{c_1, \dots, c_m\} \in C(W^*, D)$. Then there are $m!$ sequences of individuals d_1, \dots, d_m , such that the conceptual cover CC , defined by

$$CC = \{c_1 \cup \{\langle w_{n+1}, d_1 \rangle\}, \dots, c_m \cup \{\langle w_{n+1}, d_m \rangle\}\}$$

is an extension of CC^* to $C(W, D)$. Since this holds for each CC^* in $C(W^*, D)$, and since $|C(W^*, D)| = (|D|!)^{n-1}$, then there are $(|D|!)^{n-1} \times |D|! = (|D|!)^n = (|D|!)^{|W|-1}$ in $C(W, D)$. \square

¹I am indebted to Rosella Gennari and Paul Dekker for the proof of this result.

The number of possible covers grows exponentially to the number of possibilities taken into consideration. The smaller a set of worlds, the less conceptual covers are available. In an epistemic perspective, this means that the number of non equivalent ways of identifying the objects in one domain is relative to the number of possibilities compatible with the relevant information state. The more informed a state, the less identification methods are available. A corollary of proposition 4.1.3 is that with respect to a singleton set of worlds there is only one possible conceptual cover.

4.1.4. COROLLARY. If $|W| = 1$, then $|C(W, D)| = 1$ for all D .

A singleton set of worlds represents the belief state of a subject who has a complete picture of what is the case, a subject without doubts about individuals and their properties. With respect to such a state, it is irrelevant which conceptual perspective you assume, because all different ways of specifying objects collapse into one. The availability of a number of non-equivalent methods of identification is a sign of lack of information.

The relativity result shows that identification methods are not given once and for all, but depend on a set of possible worlds. Indeed, as we have already seen numerous times in the previous chapters, sets of concepts which constitute proper conceptualizations with respect to a set of possibilities may cease to do so with respect to a larger set. This is due to the fact that if we add a world, two concepts, which had distinct values with respect to all old worlds, can overlap in the new world, and so can no longer be part of one and the same conceptualization.

As an illustration, consider again the card situation above. In front of you lie two cards turned over. One is the Ace of Hearts, the other is the Ace of Spades. You don't know which is which. Furthermore, the card on the left is marked and you know it. We can model your information state σ as follows:

$$\begin{array}{lcl} w_1 & \mapsto & \heartsuit \bullet \spadesuit \\ w_2 & \mapsto & \spadesuit \bullet \heartsuit \end{array}$$

There are two conceptual covers based on σ :

$$A = \{\lambda w[left]_w, \lambda w[right]_w\}$$

$$B = \{\lambda w[Spades]_w, \lambda w[Hearts]_w\}$$

A identifies the cards by their position on the table and B identifies the cards by their suit. Which of these two methods of identification is operative, can influence evaluation. Consider, for instance, the following three examples:

(186) Which card is the Ace of Hearts?

$$?x_n (x_n = \heartsuit)$$

(187) Any card might be the Ace of Hearts.

$$\forall x_n \Diamond (x_n = \heartsuit)$$

(188) One of the cards is marked.

$$\exists x_n ((x_n)^\bullet)$$

If we assume cover A as value for n , then the question (186) partitions σ in two blocks (see chapter 1), the sentence (187) is consistent (see chapter 3), and the existential sentence (188) is supported by the state, so you are licensed to utter it (see again chapter 3). On the other hand, under a perspective which assigns the cover B to n , (186) is vacuous, (187) is inconsistent, and (188) is not supported by σ .

Suppose now you learn that the card on the right is the Ace of Hearts. Your information states will now look like the following, call it τ :

$$w_2 \mapsto \spadesuit^\bullet \heartsuit$$

In such a situation, the two identification methods by suit and by position collapse into one. A and B restricted to w_2 are one and the same set, namely the set consisting of the following two functions from $\{w_2\}$ to $\{\spadesuit, \heartsuit\}$: $(w_2 \rightarrow \spadesuit)$ and $(w_2 \rightarrow \heartsuit)$. With respect to the new state τ , choosing different conceptual perspectives does not change the update effect of the three sentences above. Under any perspective, (186) is vacuous, (187) is inconsistent and (188) is supported in τ . The first two sentences are not acceptable, because both are intended to express gaps of information, and in a state of maximal information such as τ there are no such gaps. The existential sentence (188) is instead supported irrespective of which value is assigned to n . While in the previous situation, if you had chosen the wrong method of identification, you could have failed to be licensed to utter (188), – the sentence was not supported in σ under the identification by suit –, here the correctness of your utterance is not relative to the ways of specifying the cards in the domain. In a state of maximal information, you can quantify directly over the individuals disregarding the ways in which these are identified.

Suppose now Ralph enters the room and he knows that the card on the left is marked, but he believes that the Ace of Hearts is on the left, instead of the Ace of Spades. The following diagram visualizes Ralph's information state:

$$w_1 \mapsto \heartsuit^\bullet \spadesuit$$

Suppose now you utter the following *de re* sentence:

(189) Ralph believes the Ace of Spades to be marked.

$$\exists x_n (x_n = \spadesuit \wedge \Box_r (x_n)^\bullet)$$

As soon as another subject enters the picture, conceptual covers are no longer irrelevant. The evaluation of (189) in w_2 , which from your perspective stands for the actual world, involves taking into consideration also world w_1 , which is the only possibility in Ralph's belief state (see chapter 2). With respect to $\{w_1, w_2\}$, A and B are not one and the same set of concepts. The two methods of identification by position and by suit are no longer interchangeable. Indeed, if n is assigned A, (189) is true; if n is assigned B, (189) is false.

To conclude, in a one agent situation, the presence of different contrasting perspectives over one domain is a sign of ignorance. The process of information increase can be characterized as a process of identification of different conceptualizations. In a state of maximal information all different perspectives coincide and we can talk directly about the individuals in such a state, independently from the ways in which these are specified. However, as soon as another agent enters the picture, different ways of identifying the object in the domain may arise and conceptual covers become relevant again to relate the individuals figuring in the different agents' epistemic alternatives.

4.2 Trans-world Identification

Methods of trans-world identification are 'ways of understanding questions as to whether an individual figuring in one possible world is or is not identical with an individual figuring in another possible world.'² In this section, I investigate which view of trans-world identification is formalized by the notion of a conceptual cover and I compare it with other views that have been proposed in the literature.

The most general way to define a method of cross-world identification is as a relation between world-individual pairs.

4.2.1. DEFINITION. Let W be a set of possible worlds and D a set of individuals. A *method of cross-identification* R for $\langle W, D \rangle$ is defined as follows:

$$R \subseteq (W \times D)^2$$

A method of cross-identification tells you which individual is which across the boundaries of different possible worlds. We write $\langle w, d \rangle R \langle w', d' \rangle$ to indicate that d in w is identified with d' in w' , or that d in w is the counterpart of d' in w' .

A typical example of a method of cross-identification is the following:

$$R_D = \{ \langle w, d \rangle, \langle w', d' \rangle \mid d = d' \ \& \ w, w' \in W \}$$

R_D represents the view of trans-world identification presupposed by the G&S logic of question (see chapter 1), classical Modal Predicate Logic (see chapter 2)

²Hintikka (1969), p. 33.

and Slicing (see chapter 3). Two individuals are cross-identified iff they are one and the same individual. I call this view, the *rigid* method of cross-identification.

Any set of individual concepts determines a method of cross-identification. Given a set $S \subseteq D^W$, we can define the corresponding R_S as follows:

$$\langle w, d \rangle R_S \langle w', d' \rangle \text{ iff } \exists c \in S : c(w) = d \ \& \ c(w') = d'$$

By such a construction, the set of all concepts $IC = D^W$ determines a cross-identification method which identifies all individuals in one world with all others in each other world.

$$R_{IC} = \{ \langle w, d \rangle, \langle w', d' \rangle \mid w = w' \rightarrow d = d' \}$$

R_{IC} is the method of cross-identification assumed by the first contingent identity semantics CIA in chapter 2. Any two individuals figuring in two different worlds are cross-identified. In this form, it strikes us for its intuitive inadequacy.

Conceptual covers are sets of concepts in which no splitting or merging is allowed. Each individual in one world is identified with one and only one individual in each other world. Clearly, also a conceptual cover determines a method of cross-identification. The converse obviously does not hold, not all methods of cross-identification determine conceptual covers. In this section, I am interested in specifying in a precise way the class of cross-identification methods R which do correspond to a conceptual cover. It turns out that a method of cross-identification R has to satisfy the most stringent conditions in order to determine a conceptual cover, namely:

(i) R is an equivalence relation:

- (a) R is reflexive;
- (b) R is symmetric;
- (c) R is transitive.

(ii) Each individual has one and only counterpart in each world:

- (a) $\forall w, w', d : \exists d' : \langle w, d \rangle R \langle w', d' \rangle$;
- (b) $\forall w, w', d, d', d'' : \langle w, d \rangle R \langle w', d' \rangle \ \& \ \langle w, d \rangle R \langle w', d'' \rangle \Rightarrow d' = d''$.

I will call a method of cross-identification *proper*, if it satisfies (i) and (ii). The rigid method R_D above is a typical example of a proper R . Cross-identification methods R_S determined by arbitrary sets of concepts, are symmetric, but need not be proper. R_{IC} is an example of a non-proper method of cross-identification. Although R_{IC} is reflexive, symmetric and satisfies (iia), it does not satisfy transitivity and condition (iib). The methods of cross-identification A and B below are examples of relations satisfying condition (i) and (iia), but not condition (iib), and of relations satisfying (i) and (iib), but not (iia), respectively.

$$A = (W \times D)^2$$

$$B = \{\langle w, d \rangle, \langle w, d \rangle \mid d \in D \ \& \ w \in W\}$$

In A , all individuals in all worlds are identified. In B , no two individuals from two different worlds are identified.

Finally, to see that conditions (i) and (ii) are independent consider the counterintuitive method C which satisfies (ii), but is not an equivalence relation. Let $|W| = |D| = 2$:

$$C = \{\langle w, d \rangle, \langle w', d' \rangle \mid d \neq d'\}$$

I will show now that a method of cross-identification R corresponds to a conceptual cover iff R is proper. For more detailed proofs see Appendix A.4.

4.2.2. PROPOSITION. Let CC be a conceptual cover over (W, D) . The method of cross-identification R_{CC} determined by CC is proper.

Recall that R_{CC} is defined so that $\langle w, d \rangle R_{CC} \langle w', d' \rangle$ iff $\exists c \in CC : c(w) = d$ and $c(w') = d'$. The existence condition on conceptual covers implies that reflexivity and condition (iia) are satisfied (together with the fact that we are dealing with total functions). The uniqueness condition on CC s implies transitivity and condition (iib). Symmetry follows by construction of R_{CC} .

Let's see now how we can define a conceptual cover from a proper method of cross-identification.

4.2.3. DEFINITION. Let R be a cross-identification method over (W, D) . The set of classes of pairs induced by R is the following set:

$$CP_R = \{[w, d]_R \mid w \in W \ \& \ d \in D\}$$

where $[w, d]_R = \{\langle w', d' \rangle \mid \langle w, d \rangle R \langle w', d' \rangle\}$.

4.2.4. PROPOSITION. The set of classes of pairs induced by a proper cross-identification method R is a conceptual cover.

This result follows directly from the two lemmas below.

4.2.5. LEMMA. Let R be a proper method of cross-identification over (W, D) . Then

$$\forall \alpha \in CP_R : \forall w \in W : \exists! d \in D : \langle w, d \rangle \in \alpha$$

Lemma 4.2.5 states that if R is proper, then each element of CP_R uniquely determines a total function from W to D , that is an individual concept.

I write $\alpha(w)$ to denote the individual d such that $\langle w, d \rangle \in \alpha$.

4.2.6. LEMMA. Let R be a proper method of cross-identification over (W, D) . Then

$$\forall w \in W : \forall d \in D : \exists! \alpha \in CP_R : \alpha(w) = d$$

The proof of this lemma follows directly from the fact that if R is an equivalence relation, then CP_R is a partition of the set of world-individual pairs. Lemma 4.2.6 states that the set of concepts CP_R satisfies the uniqueness and the existence conditions and therefore is a conceptual cover.

Each conceptual cover uniquely determines a proper method of cross-identification (proposition 4.2.2) and each proper method of cross-identification uniquely determines a conceptual cover (proposition 4.2.4).

The existence and uniqueness conditions, which characterize a conceptual cover, thus find further justification from this perspective. Condition (i) is a quite natural constraint on methods of identification. In each world, each individual should be identified to itself (reflexivity). If d in w is identified with d' in w' , then d' in w' must be identified with d in w (symmetry). Finally, if d in w is identified with d' in w' and d' in w' is identified with d'' in w'' , then also d in w and d'' in w'' must be identified (transitivity).

Condition (ii) is also an intuitive constraint. On the one hand, it simply says that two individuals cannot become one. On the other, it requires that no individual can cease to exist once we move from one world to the other.

The rigid view of trans-world identification exemplified by R_D above, being proper, corresponds to a conceptual cover. However, many other methods of identification are proper as well. In the analysis I defend in this thesis, different proper methods are allowed to act as the operative ones in different occasions.

4.2.1 Alternative Views of Trans-world Identification

After having seen which view of trans-world identification is formalized by the notion of a conceptual cover, I will now review a number of alternative views that have been proposed in the literature. The comparison will also give us the occasion to discuss a series of future applications and loose ends of the present analysis.

Individuating Functions

The present thesis can be seen as a development of a simple insight that Jaakko Hintikka presented in two articles at the end of the 60s where he envisaged the availability of different methods of cross-identification on different occasions.³ In these articles, however, Hintikka did not carry his own insight far enough. The empirical applications he discusses are not totally convincing and the formalization he assumes is quite unsatisfactory. Hintikka (1969) discusses the logic of

³See Hintikka (1967) and Hintikka (1969).

propositional attitudes. Hintikka recognizes that quantification in such a logic presupposes a method of cross-identification and formalizes such methods by sets F of (possibly partial) individuating functions f mapping worlds to individuals, satisfying the following non-splitting condition: $\forall w : \forall f, f' \in F : (f(w) = f'(w) \Rightarrow \forall w' : f(w') = f'(w'))$, which corresponds to my uniqueness condition. Although he implicitly proposes to relativize quantification to such methods of identification, Hintikka does not discuss empirical applications of such a relativization in connection with propositional attitude reports. For an example of such a concrete application, we have to go back to Hintikka (1967), where he discusses the semantics of perception reports. In this article, Hintikka isolates two distinct methods of cross-identification (perspectival or demonstrative cross-identification vs physical cross-identification) and introduces two kinds of quantifiers $\exists x$ and Ex , the former ranging over individuals identified by the first method, the latter ranging over individuals identified by the second method. This distinction is used to capture the contrast between direct and indirect perception reports of the following kind:

(190) d perceives Mr. Smith.

$\exists x(d \text{ perceives that Mr. Smith} = x)$

(191) d sees who the man in front of him is.

$Ex(d \text{ perceives that the man in front of } d = x)$

I will not discuss Hintikka's logic of perception that has been criticized (see Barwise and Perry (1983) for a more influential analysis of perception reports). Hintikka's insight that perceptive reports and propositional attitude reports can involve different methods of cross-identification is very impressive, but it should be generalized substantially. Hintikka's strategy to introduce a different quantifier for each method of cross-identification is not satisfactory. Obviously, far more than two methods are operative in our ordinary conversations, and so, such a strategy would lead to an intolerable complication of the syntax, while not explaining the context sensitivity involved in these cases.

Other authors have proposed a semantics involving quantification over sets of individual concepts, for instance Kraut (1983) and Zeevat (1995). I will just consider the former. Kraut (1983) discusses propositional attitude reports and proposes to analyze them in terms of quantification over a set of individuating functions but avoids Hintikka syntactic complications by making the domain of quantification explicitly context dependent. Kraut's approach constitutes a clear example of what I called a pragmatic analysis in chapter 2. Although they are close in spirit, the theory I defend in chapter 2 and Kraut's analysis depart on a series of points. Firstly, Kraut does not discuss any particular extra constraint on his sets of individuating functions (in particular no uniqueness condition is assumed) and, in addition, his analysis of *de dicto* attitude reports is quite different

from mine. As announced in the title of the article, according to Kraut, there are no *de dicto* attitudes. Kraut proposes the ‘*de re*’ representation in (193) for the traditional *de dicto* reading of a sentence like (192):

(192) Ralph believes that the tallest member of the club is a spy.

(193) $\exists x \Box_r(x \text{ is the tallest member of the club} \ \& \ x \text{ is a spy})$

This means that on Kraut’s account, traditional *de dicto* reports show the same context dependence as traditional *de re* reports and this seems to me at variance with our intuition. Consider a situation in which Ralph would assent to the sentence ‘The tallest member of the club is a spy’ while having no idea as to the identity of the tallest member of the club. According to the analysis in chapter 2, the *de dicto* reading of (192), which is represented in accordance with the tradition, by sentence (194), is true in such a situation irrespective of the operative method of identification.

(194) $\Box_r \exists x_n(x_n \text{ is the tallest member of the club} \ \& \ x_n \text{ is a spy})$

On the other hand, on Kraut’s account, the acceptability of (192) is relative to the conceptual perspective we assume. Indeed, in order for (193) to be true, the concept ‘the tallest member of the club’ must be part of the operative method of identification. Now, such a method is clearly not the prominent one in the described circumstances (Ralph has no idea who the tallest person in the club is, it is said), so a shift of identification method would be required in order to interpret such a simple and ordinary case. I don’t think this is correct. Not all attitude reports are relative to a method of identification and the so-called *de dicto* readings are precisely those for which there is no such dependence.

The reason why Kraut assumes such an analysis of *de dicto* reports has to do with his main empirical application, namely intentional identity phenomena. In order to be able to represent Hob-Nob sentences like (195a) by a ‘geachean’ representation (195b), the existential quantifier must take wide scope also in traditional *de dicto* sentences:

(195) a. Hob thinks that a witch has blighted Bob’s mare, and Nob believes that she [the same witch] killed Cob’s sow.

b. $\exists x(\Box_h \phi \wedge \Box_n \psi)$

Kraut’s solution to the intentional identity puzzle is not totally satisfactory though. For instance, if we can represent (195a) by (195b), then (196b) is also a possible representation for (196a), but then we lack an explanation of the unavailability of the anaphoric relation in the latter case:⁴

⁴Kraut might argue in favor of a pragmatic solution to this difficulty which involves the use of partial concepts. By assuming a method of identification containing a concept – the relevant

(196) a. Hob thinks that a witch has blighted Bob's mare. (?) She killed Cob's sow.

b. $\exists x(\Box_h \phi \wedge \psi)$

The difficulty I have indicated is just one of the many intricacies arising from the obscure area of intentional identity. Sets of individuating functions as well as conceptual covers can be used to determine which individuals corresponds to which in different agents' states. Therefore, Hob-Nob phenomena seem a natural application for a system using individual concepts. However, things are far more complex than they seem and, therefore, I must leave a proper treatment of these phenomena to another occasion.

Counterpart Relation

Lewis (1968) introduces the counterpart relation as a substitute for identity between things in different worlds. Lewis is discussing a metaphysical notion and not an epistemic or doxastic one. Metaphysical considerations lead him to assume disjoint domains, so that no individual is allowed to figure in two different worlds. Lewis' counterpart relation C is a relation on $\bigcup_{w \in W} D_w$, where dCd' means that d is the counterpart of d' in another world. Lewis assumes that C must satisfy the constraint that if d and d' are both in D_w , then dCd' iff $d = d'$. The counterpart relation is intuitively based on the relation of similarity between individuals figuring in different worlds which need be neither symmetric nor transitive.⁵ On the other hand, each individual is similar to itself, so C is assumed to be reflexive. In our formulation, we can express Lewis counterpart relation by means of the following conditions on the method of cross-identification R :

1. $\forall w, d, d' : \langle w, d \rangle R \langle w, d' \rangle \Rightarrow d = d'$;
2. $\forall d : \exists w : \langle w, d \rangle R \langle w, d \rangle$;
3. $\forall w, w', d : \langle w, d \rangle R \langle w', d \rangle \Rightarrow w = w'$.

Condition 1 is Lewis' constraint mentioned above that in one and the same world, an individual can only have one counterpart, namely itself. Condition 2 corresponds to the constraint that C is reflexive. The last condition expresses the fact that no individual can occur in two different worlds (we can define the set of individuals D_w figuring exclusively in w as follows: $D_w = \{d \in D \mid \langle w, d \rangle R \langle w, d \rangle\}$). If we let R satisfy conditions 1, 2, and 3, we then obtain that dCd' iff $\exists w, w' :$

witch – which is defined in all Hob's and Nob's possibilities, but undefined in the actual world, we can account for the acceptability of (195a) and the unacceptability of (196a). However, the contrast between the two sentences seems to me to be more structural than a question of pragmatics. Opinions about this issue may diverge though.

⁵Lewis (1968), pp. 115-116.

$\langle w, d \rangle R \langle w', d' \rangle$. Notice that conditions 1 and 2 are satisfied also by conceptual covers (condition 2 is reflexivity and condition 1 follows from reflexivity and condition (iib)). Condition 3 is instead not needed if we assume an epistemic perspective. Conceptual covers further satisfy symmetry, transitivity and the condition that each individual has one and only one counterpart in each world, for which I have already argued for. Counterpart relations have also been assumed in connection to epistemic or doxastic phenomena by Lewis himself in (1983), but also in Stalnaker (1987) and van Rooy (1997). In the latter, counterpart theory is used to account for the double vision puzzles of *de re* belief attributions. Van Rooy proposes to analyze these constructions by means of quantification over sets of counterpart functions.⁶ The analysis I defend in chapter 2 and van Rooy's semantics of belief reports have the same empirical coverage, at least as far as double vision cases are concerned.⁷ On the other hand, from a formal point of view, the two analyses are quite different, and mine is more conservative in staying as close as possible to ordinary modal predicate logic. This feature has many advantages in terms of simplicity and tractability of the logic, e.g. the empirical predictions of the system are easily seen and its formal properties, notably the completeness result, can be proved by standard techniques. This is not always the case in van Rooy's system.

Individuation Schemes

In two short articles,⁸ Jelle Gerbrandy presents, in a lucid way, the central idea I defend and investigate in this thesis: different methods of identification are operative in different contexts and evaluation of fragments of the discourse can vary relative to these methods. Gerbrandy also discusses the question of cross-identification in connection with epistemic phenomena, in particular knowing-who constructions, and proposes to relativize their interpretation to what he calls *individuation schemes*. These are methods of cross-identification R where R is an equivalence relation.

In chapter 1, section 1.4.3, I have already argued in favor of the adoption of

⁶Van Rooy's counterpart functions C map individual-world pairs (d, w) to individuals d' element of D_w . By adopting counterpart functions rather than relations he rules out the possibility of one individual having two counterparts in some world (see condition (iib) above). Van Rooy can still express double vision cases, which crucially involve such splitting, by taking sets of possibly 'overlapping' counterpart functions rather than single counterpart relations.

⁷I am not sure about the predictions of van Rooy's analysis in relation to the other problems I have discussed in chapter 2, e.g. the shortest spy problems, but I guess my analysis does slightly better in connection with these cases. Although van Rooy recognizes the context sensitivity of belief attributions, this is not reflected in his semantics where he seems to follow Kaplan and encodes information about which counterpart functions are suitable for quantification in the model (see his notion of counterpart functions by acquaintance C_{acq}). But, as I argued in chapter 2, such a strategy cannot solve the shortest spy problems without generating other difficulties (see the cases of Odette's lover and Susan's mother in chapter 2, section 2.3.2).

⁸Gerbrandy (1997) and Gerbrandy (2000).

more stringent constraints⁹ for R in connection with questions and knowing-who constructions. My argument there had to do with the intuitive relation between knowing-who and knowing-how many constructions. If we relativize questions to possibly non-proper identification methods, we predict that one can know who is P , without knowing how many are P . The following sentence would be consistent, and this would be at variance with our intuition:

- (197) I don't know how many people were late today, but I know who was late today.

It could be argued, however, that there are interpretations under which this sentence might be acceptable, namely, if we interpret the second conjunct as saying that you know which *kinds* of individuals were late today, rather than which *individuals*. For instance, you could perhaps consistently say:

- (198) I don't know how many were late today, but I know who was late today, namely some linguists and some logicians.

A possible justification for dropping condition (ii), which is otherwise quite intuitive, is that by taking equivalent cross-identification methods R in which two or more individuals can become one, we can account for this individual-kind ambiguity. However, the individual-kind distinction seems to be of a different nature than the one formalized by different covers. While the latter involves different (though equivalently fine-grained) ways of identifying the entities in the domain, the former involves looking at one domain assuming different levels of granularity. Gerbrandy's individuation schemes might be used to formalize such differently fine-grained ways of conceiving one domain. Ginzburg (1995) accounts for examples like (198) by adopting a relative notion of answerhood.¹⁰ By assuming that different contexts can select different individuation schemes as quantificational domains for the *wh*-phrases, we might be able to deal with the individual-kind distinction while maintaining an ordinary notion of answerhood. However, spelling out the details of such an analysis is not a trivial task. Furthermore, it is not totally clear to me whether a domain selection strategy is the correct one here, since there is a lot of vagueness playing a role, and therefore an analysis like that in Ginzburg (1995) might be more appropriate for these cases.

4.3 Towards a Pragmatic Analysis

The main idea of the analysis I defend in this thesis is that different methods of identification are available on different occasions and that evaluation of sentences

⁹Namely R should correspond to a conceptual cover, that is, it should be a proper method of cross-identification, i.e, it should also satisfy condition (ii) and not only be an equivalence relation.

¹⁰See chapter 1, section 1.5.3 for a discussion of Ginzburg's analysis.

can vary relative to these methods. I have proposed to characterize methods of identification by means of the notion of a conceptual cover and I have studied a number of linguistic constructions whose interpretation depends on the contextually selected conceptualization, notably questions and knowing-who constructions (chapter 1), belief attributions (chapter 2), epistemic modals, presupposition and specific indefinite NPs (chapter 3). Such constructions can express different meanings in different contexts. The question I will explore in the remaining part of this chapter is how the addressee may be able to select the intended identification method while interpreting these constructions. In order to shed some light on this complicated issue, I will use notions from Optimality Theory (OT) and Game Theory (GT). The analysis I propose in the following pages is still in its germinal phase, and needs further investigation. Nevertheless, it shows interesting aspects of the cover selection procedure, the most significant one being that shifts of cover never occur without justification. Furthermore, it illustrates that the use of OT and GT notions in the explanation of phenomena lying on the semantics-pragmatics interface is promising, although not totally unproblematic.

According to an OT analysis of interpretation (see de Hoop and Hendriks (1999)) the process of interpretation of natural language sentences is ruled by a number of ordered interpretation constraints. The addressee chooses from a set of possible meanings the ones which optimally satisfy these constraints. A simple machinery of this sort is sufficient to give a rational explanation of the cover selection procedures involved in many of the examples we have considered in the previous chapters, but not all. A number of potentially problematic cases of *de re* sentences cannot be explained by such an addressee-oriented analysis. I will suggest that a proper treatment of such examples requires a bi-dimensional interpretation theory (see Blutner (1999) and Blutner and Jäger (1999)), in which also the speaker's perspective is taken into consideration. In such a theory, the optimal solution is searched along two dimensions, the one of the addressee and the one of the speaker whose choice of uttering this or that sentence is influenced firstly by general principles of generation, secondly by the principle of cooperation, and finally by her particular interests and goals. I will follow Dekker and van Rooy (1999) and recast bi-dimensional OT interpretation processes in terms of 'interpretation games'. Game Theory turns out to be a promising framework for describing the interplay of general linguistic constraints and particular goals in the search for an optimal interpretation.

4.3.1 OT Interpretation Theory

In Optimality Theory (see Prince and Smolensky (1997)) conflicts between constraints are arbitrated by ranking one constraint over the other. OT has been applied in phonology, where it constitutes the dominant theoretical paradigm, in syntax, and, recently, also in semantics and at the semantics-pragmatics interface (see de Hoop and de Swart (1999), de Hoop and Hendriks (1999), Blutner (1999),

Blutner and Jäger (1999), Zeevat (1999a), Zeevat (1999c) and Dekker and van Rooy (1999).)

An OT interpretation theory is based on a set of constraints ordered according to their relative strength, which help us in deciding between different readings allowed by the generative part of the grammar. The addressee has a set of alternative contents for a specific expression at her disposal. The best interpretations are those elements of the set which do better on the interpretation constraints than all other alternative candidates, where candidates that have arbitrary many violations of lower ranked constraints do better than candidates that have also one violation of a higher ranked constraint.

The phenomena that we have considered in the previous chapters provide evidence in favor of competition and ranking of interpretation constraints. In the following pages, I will briefly discuss a number of constraints which seem to play a role in the process of cover selection. I will then present examples of conflicts between these constraints and I will discuss which ranking is suggested by these conflicts. In the discussion, I follow Blutner (1999) who adopts updates in an OT setting. Meanings are identified with information change potentials (see chapter 3). Information states are formalized as sets of possibilities. Assertions modify these states in various ways by eliminating possibilities or extending them. Questions are taken to partition the states in alternative blocks.¹¹ As in the previous chapters, sensitivity to methods of identification is expressed by means of *CC*-indices n, m, \dots . I will write $UP_{\wp}^{\varphi}(\alpha[\vec{x}_{\vec{n}}])$ to denote the set of the potential outcomes of updating a state σ with an expression $\alpha[\vec{x}_{\vec{n}}]$ under perspective \wp . As in chapters 1 and 3, a conceptual perspective \wp is a function assigning conceptual covers to the *CC*-indices \vec{n} occurring in α .

Interpretation Constraints

The first interpretation constraint I will discuss is the principle of ANCHOR discussed in Zeevat (1999a).

ANCHOR says that interpretation should be anchored to the context. This principle governs the interpretation of expressions which are assigned a value either by deixis or by anaphora resolution, and hence should find a proper antecedent in the context. Examples of such expressions are pronouns, tenses and *CC*-indices. Normally antecedents for such expressions are made salient either by explicit mention in the preceding discourse (anaphora resolution) or by the actual presence of the relevant referents in the utterance situation (deixis). *CC*-indices are no exception to this. Consider again the Spiderman, Ortcutt and butler examples that have been presented in chapter 1, 2 and 3 respectively. Recall the relevant situations:

¹¹See Groenendijk (1998) and Groenendijk (1999) for a dynamic treatment of interrogative sentences.

Spiderman Someone killed Spiderman. You have just discovered that John Smith is the culprit. So you can say (a). Now John Smith is attending a (masked) ball. You go to arrest him there, but you don't know what he looks like. So you say (b).

- (199) a. (John Smith did it. So) I know who_n killed Spiderman.
 b. (This person might be the culprit. That person might be the culprit.
 So) I don't know who_m killed Spiderman.

Ortcutt You can tell each half of the Orcutt story separately. In one half Ralph sees Orcutt wearing the brown hat. In the other he sees him on the beach. From the first story you can reason as in (a). From the second story you can reason as in (b).

- (200) a. Ralph believes that the man with the brown hat is a spy.
 The man with the brown hat is Orcutt.
 So Ralph believes of Orcutt that he_n is a spy.
 b. Ralph believes that the man seen on the beach is not a spy.
 The man seen on the beach is Orcutt.
 So Ralph does *not* believe of Orcutt that he_m is a spy.

the butler Suppose a butler and a gardener are sitting in a room. One is called Alfred and the other Bill. We don't know who is who. The butler is the culprit.

- (201) a. The butler did it. So it is not true that anybody $_n$ in the room might be innocent.
 b. Alfred might be innocent. Bill might be innocent. So anybody $_m$ in the room might be innocent.

In the (b) case of the Spiderman example, the value for the index m is suggested by the concepts given by the visual images of the masked faces, which become salient by entering the perceptual field of the participants in the conversation. In all other cases, one conceptualization or the other is suggested as value for the relevant index by the previous discourse which explicitly mentions one or the other concept. In all examples, the context supplies as antecedents for the *CC*-indices single isolated concepts rather than the conceptual cover themselves. In contrast with the case of pronouns, it seems that the context can contribute to determine the value of a *CC*-index n by merely suggesting conditions for this value rather than by supplying the value itself. These conditions have normally the following form:

$$\{c_1, \dots, c_m\} \subseteq n$$

where c_1, \dots, c_m are concepts salient in the context. Such conditions can clearly fail to uniquely determine a specific conceptual cover. In these cases, the content expressed by a sentence $\phi[x_n]$ can fail to be uniquely determined. By the specification of a set of possible conceptualizations, a set of possible contents can be selected, rather than one. This indeterminacy, though, does not necessarily lead to failure of communication. The condition for avoiding such a failure is that all these contents behave uniformly with respect to the relevant background state. The point of an assertion is to reduce the background state in a certain determinate way. If the sentence is associated with a number of contents that affect the input state in different ways, it would be unclear whether a possibility should be eliminated or included in the resulting state. On the other hand, if all the contents involve the same action on the background state, no ambiguity arise (see Stalnaker (1978)). Therefore, we can assume that a condition restricting the possible values of a *CC*-index n that uniquely determines the effect of the utterance of a certain sentence $\phi[x_n]$ on the input state can act as a suitable antecedent for n . More formally, if for all perspectives \wp, \wp' the following holds: $\{c_1, \dots, c_m\} \subseteq \wp(n) \ \& \ \{c_1, \dots, c_m\} \subseteq \wp'(n) \Rightarrow UP_\sigma^\wp(\phi(n)) = UP_\sigma^{\wp'}(\phi(n))$, then the sequence c_1, \dots, c_m can act as a suitable antecedent for n . In a context in which such a sequence of concepts is salient ANCHOR can be satisfied. As an illustration, consider the final *de re* sentence in the (a) case of the Ortcutt example:

(202) Ralph believes of Ortcutt that he_n is a spy.

$$\exists x_n(x_n = o \wedge \Box S(x_n))$$

Let σ be any state resulting from an update with the two preceding sentences in (200a). The condition in (203) is sufficient for a felicitous interpretation of (202) in such a situation, because, in σ , the update brought about by (202) will be the same under any cover which satisfies (203):

(203) $\lambda w[\text{the man with the brown hat}]_w \in n$.

A condition like the following might instead cause indeterminacy:

(204) $\lambda w[\text{the shortest spy}]_w \in n$.

Since it may not be supported by the background state σ that Ortcutt is the shortest spy, it would not be clear under which perspective to identify Ortcutt in Ralph's belief's state, so (202) can turn out both supported and rejected in σ .

Now, ANCHOR says that interpretation should be anchored. All anaphoric expressions should find a proper antecedent in the context and we have just discussed the peculiar modalities of the anchoring of *CC*-indices. Still assuming that the absence of a suitable salient antecedent for a *CC*-index leads to communication breakdown is quite unrealistic. In real life communication, people deal with these cases by accommodating one or the other (condition on) conceptualization.

So, in such situations we should allow accommodation. However, accommodation should be disallowed in case a proper antecedent is already available in the context. OT interpretation theory can capture the latter intuition by assuming a principle which prohibits the addressee the addition of new material to the context, Zeevat (1999a) calls such a principle ***ACCOMMODATION** (see also Blutner (1999)). But, OT can also account for the fact that accommodation is allowed in certain circumstances by positing that ANCHOR can overrule ***ACCOMMODATION** (see again Zeevat (1999a)). If no antecedent is available we choose to accommodate in order to satisfy ANCHOR which ranks higher than ***ACCOMMODATION**. If an antecedent is already present in the context, ANCHOR is satisfied. Consequently ***ACCOMMODATION** is the critical constraint and its violation become crucial. So, we prefer readings which do not involve accommodation.

ANCHOR in interaction with ***ACCOMMODATION** explains the contrast between the (a) and the (b) cases in the Spiderman, Orcutt and butler examples. I will just consider the latter case:

- (205) a. (The butler did it. So) it is not true that anybody_{*n*} in the room might be innocent.

$$\neg \forall x_n \Diamond I(x_n)$$

- b. (Alfred might be innocent. Bill might be innocent. So) anybody_{*m*} in the room might be innocent.

$$\forall x_m \Diamond I(x_m)$$

Consider example (205b). The *CC*-index *m* should intuitively be assigned the cover $A = \{\text{Alfred, Bill}\}$ rather than the cover $B = \{\text{the butler, the gardener}\}$. By assuming ANCHOR and ***ACCOMMODATION** we explain this preference. *A* is salient in the context, whereas *B* is not. Hence, an assignment of *m* to *B* would violate ***ACCOMMODATION**, whereas an assignment of *A* to *m* would not involve any violation. By the same kind of reasoning, *n* is assigned cover *B*, and not *A* in (205a).

Another clear example in which ANCHOR plays a crucial role is the following case discussed in chapter 1.

Priscilla Consider sentence (206) uttered by Priscilla in the two situations α and β :

α : In your living room.

β : At a party with many African leaders.

(206) Who is the president of Mali?

- a. Konare is the president of Mali.
- b. He [pointing at Konare himself] is the president of Mali.

On the one hand, an interpretation of the question under a demonstrative cover containing Konare himself is clearly not anchored in context α and accommodating in this case would mean flying to Africa and kidnapping Konare. On the other hand, in context β , (206) can be interpreted under the demonstrative cover without violations. Indeed, (b) is a good answer to (206) in β , but not in α .

The next principle I will discuss is a constraint which specifically governs cover selection procedures.

***SHIFT** expresses a general preference for interpretations which do not involve shift of conceptualizations. As an illustration of the role of *SHIFT, consider the following variation on the workshop example discussed in chapter 1:

the workshop You are attending a workshop. In front of you lies the list of names of all participants, around you are sitting the participants in flesh and blood. You don't know who is who. Consider now the following question:

(207) Who_{*n*} has taken whom_{*m*} to the party yesterday night?

$$?x_n y_m R(x_n, y_m)$$

Although two conceptualizations are salient in the described situation, namely naming and the ostensive cover, there is a clear preference for a uniform interpretation for *n* and *m*. Indeed, in such a situation, where you do not know which person is called what, replies like (208) or (209) are intuitively more acceptable answers to (207), than a reply like (210):

(208) Dylan Dog has taken Nathan Never. Ken Parker has taken Dylan Dog. ...

(209) This man has taken that man. The man in the first row has taken that guy over there. ...

(210) This man has taken Nathan Never. That man has taken Ken Parker. ...

Before turning to cases in which *SHIFT is overruled, I will discuss the principle of STRENGTH.

STRENGTH is a constraint which specifically governs the process of cover selection and expresses a preference for the selection of stronger covers.¹² Intuitively, if A is a stronger cover than B, then A represents the objects of our experience better than B. Before trying to give a formal characterization of the notion of a stronger cover, I will discuss a somewhat artificial, but clarifying example, of which the interpretation seems to be ruled by such a constraint.

Consider the following situation. Suppose you have two neighbors, the upstairs neighbor who is called John Smith and who is the major of your town and the downstairs neighbor who is called Bill White and who is the local mafia boss. You also have two roommates K and J. Suppose now that (a) K knows that John Smith is the major, and Bill White is the local mafia boss, but he has no idea that they are living in your own apartment block; (b) J knows that the major is living upstairs and the mafia boss downstairs but he has no idea what they are called.

Suppose now that John Smith has called your apartment for one or other reason. K and J come home at different times and ask you separately:

(211) Who_n called?

The intuition is that not the same cover can be accommodated as value for *n* on the two occasions. For instance, you could accommodate naming as value for *n* in case K asked the question, but not in the other case. Indeed,

(212) K: Who_n called?

- a. John Smith.
- b. (?) The upstairs neighbor.
- c. The major.

(213) J: Who_n called?

- a. (?) John Smith.
- b. The upstairs neighbor.
- c. The major.

Let A be naming, B the cover containing the concepts ‘the upstairs neighbor’ and ‘the downstairs neighbor’, and C the third relevant cover containing the concepts ‘the major’ and ‘the local mafia boss’. Now I will say that A and C are stronger than B in the first case, whereas B and C are stronger than A in the second

¹²This principle is different from Zeevat’s principle of strength which expresses preference for informationally stronger readings. As will become clear soon, if A is stronger than B, an interpretation under cover A does not have to be informationally stronger than an interpretation under cover B.

case. This ranking is predicted by the definition of relative strength of a cover which I propose below according to which the strength of a cover depends on the number of alternative relevant covers that have been identified with it in the specific information states of the relevant agents. By assuming a constraint like STRENGTH we can account for our intuitions with respect to this example.

The general idea of the definition of relative strength is that ‘clusters’ of covers are stronger than isolated ones. The notion of a strong cover is highly context dependent. There are no absolute strong covers. A cover is stronger than another with respect to the set of covers that we take into consideration, and relative to a(n) (set of) information state(s).

I first define the auxiliary notion of a cover restricted to a set of worlds σ , and the notion of an equivalence class of covers in a set C with respect to σ .

4.3.1. DEFINITION. Let CC be a cover based on (W, D) , and $\sigma \subseteq W$. The restriction of CC to σ , CC_σ is defined as follows:

$$CC(\sigma) = \{c \in D^\sigma \mid \exists c' \in CC : \forall w \in \sigma : c(w) = c'(w)\}$$

4.3.2. DEFINITION. Let C be a set of covers based on (W, D) , $CC \in C$, and $\sigma \subseteq W$.

$$[CC]_{\sigma, C} = \{CC' \in C \mid CC(\sigma) = CC'(\sigma)\}$$

I define now the relation being at least as strong as, $\gg_{\sigma, C}$ on a set of covers C with respect to a set of worlds σ .

4.3.3. DEFINITION. [Relative Strength] Let CC_1 and CC_2 be two covers in C based on (W, D) , and $\sigma \subseteq W$.

$$CC_1 \gg_{\sigma, C} CC_2 \text{ iff } |[CC_1]_{\sigma, C}| \geq |[CC_2]_{\sigma, C}|$$

A cover which is identified with n relevant covers in a state is stronger than a cover which is identified with m relevant covers in that state if n is larger than m . In particular, a cover that has not been identified with any other is weak and normally not preferred as a domain of quantification.

What makes a concept stronger than another is not the semantic nature of the term which is used to denote it, e.g. proper names versus definite descriptions, but rather the number of alternative relevant concepts that have been identified with it in the specific information states of the relevant agents. In the example above, the concept ‘John Smith’ is stronger than the concept ‘the upstairs neighbor’ with respect to K’s state, because K has identified the former concept with ‘the major’, whereas he has not identified ‘the upstairs neighbor’ with any other relevant concept. With respect to J’s state the opposite holds. As a further illustration of the relativity of the notion of a strong concept, consider the concept of your mother. With respect to your information state such a concept is stronger than

the concept ‘the shortest spy’ (unless you are different from most of us). On the other hand, in someone else’s information state, say mine, the two concepts can instead be equally strong. This is due to the fact that in your information state the former concept is actually a *cluster* of concepts, but not in mine. Indeed, you can name your mother, you know what she looks like, and you can identify her by various means, whereas I cannot do anything of this sort. Thus, with respect to my information state, but not with respect to yours, the concept ‘the reader’s mother’ will not be stronger than ‘the shortest spy’, which is also typically not identified with any other representation.

From this example it is clear that the objects of our experience are better represented by clusters of concepts, arising from the combination of elements of different covers, rather than by single isolated concepts. Strong concepts better represent genuine objects, than weak concepts and, therefore, the former can better serve as values for our variables. This is the meaning of STRENGTH.

Synopsis We have discussed four constraints so far. I suggest the following tentative ranking:

ANCHOR > *SHIFT, *ACCOMMODATION > STRENGTH

We have already seen that ANCHOR is assumed to be harder than *ACCOMMODATION in the literature (see Zeevat (1999a)). My hypothesis, which, however, needs further test, is that STRENGTH is weaker than *ACCOMMODATION. Indeed, we normally prefer weaker salient covers over stronger non salient ones. *SHIFT and *ACCOMMODATION are hard to compare. You cannot violate *SHIFT in order to satisfy *ACCOMMODATION and you cannot violate *ACCOMMODATION in order to satisfy *SHIFT. So I assume they are not ranked in any way.¹³

Putting aside the issue of the exact ordering between these constraints, notice that, by means of them, we can account for most cases of cover selection in ordinary situations. For instance, we can explain why demonstrative covers are normally preferred over alternative descriptive covers. First of all, if a demonstrative cover is available at all, then it is salient and, hence, its use cannot violate ANCHOR or *ACCOMMODATION. Furthermore, in standard situations, demonstrative covers are usually *clusters* of covers. If you have a person in front of you, you can point at her, but you can also describe her according to different parameters (visual image of her face, seize, age, etc.) and, in many cases, you can also name her. Demonstrative covers are normally strong and, hence, they

¹³This suggests that *SHIFT and *ACCOMMODATION could maybe be formulated together in a more general constraint. Notice that both are closely related to the principle Don’t Overlook Anaphoric Possibilities (DOAP) from Williams (1997) also discussed in de Hoop and de Swart (1999) and de Hoop and Hendriks (1999), which requires to seize opportunities to anaphorize text. I leave this issue as a subject for future study.

are preferred also by STRENGTH. Note that typical cases in which demonstrative covers are not selected are cases like the man with the hood (see chapter 3), in which since the relevant people are dressed up, the demonstrative cover is crucially not identified with any good descriptive cover or with naming and, therefore, it is less strong.

The combination of the four constraints discussed above explains why in many ordinary cases we don't even notice the presence of conceptual covers and the classical notion of quantification seems to be sufficient. The addressee starts assuming the strongest contextually prominent cover or accommodates one, and just stays with it for the rest of the discourse. However we can easily find situations in which the weaker three constraints are crucially overruled and these are situations in which quantification under conceptual cover plays an essential role. I will discuss four cases that we have already encountered in the previous chapters, namely the double vision (chapter 2), the soccer game (chapter 3), the workshop (chapter 1), and the Ann-Bea (chapter 2) situations. In these examples, *SHIFT is overruled in order to avoid violations of general principles of rational conversation.¹⁴ The suggested ranking is then that the latter are harder than the former. This is in accordance with what is normally assumed in the literature (see Zeevat (1999a), and also Stalnaker (1978)). I will discuss three of such general principles of conversation: CONSISTENCY, *TRIVIAL and RELEVANCE.

CONSISTENCY is a constraint which expresses preference for interpretations that do not conflict with the context (see Grice's Maxim of Quality, Stalnaker's first principle of rational conversation in (1979), and also Zeevat (1999a) and van der Sandt (1992)). Recall the following situations discussed in chapter 2 and 3 respectively, in which CONSISTENCY plays a crucial role:

double vision Ralph ascribes contradictory properties to Ortcutt since, having met him on two quite different occasions, he is 'acquainted' with him in two different ways.

- (214) a. Ralph believes Ortcutt to be a spy; and Ralph believes Ortcutt not to be a spy.

$$\text{b. } \exists x_n(x_n = o \wedge \Box S(x_n)) \wedge \exists y_m(y_m = o \wedge \Box \neg S(y_m))$$

the soccer game Suppose you are attending a soccer game. All of the 22 players are in your perceptual field. You know their names, say a, b, c, ..., but you don't recognize any of them. Consider the following sentence:

- (215) a. Anyone might be anyone.

¹⁴See Grice's theory of conversation and Stalnaker (1978).

$$\text{b. } \forall x_n \forall y_m \Diamond (x_n = y_m)$$

In both the double vision and the soccer game examples we have two conflicting constraints. On the one hand, we have *SHIFT which suggests to interpret m as n . On the other, the fulfillment of CONSISTENCY prevents this resolution, because it would render the sentences inconsistent with the context. In the double vision case, this resolution leads to inconsistency, because the sentence would say that Ralph's beliefs are contradictory, but we know that this is not the case as Ralph is not logically insane, since he simply lacks some information. In the soccer game, if n and m are assigned one and the same value the sentence is false in the described situation in which the relevant domain is not a singleton. Intuitively, an assignment of different values for n and m is what is normally assumed by an interpreter of such sentences, this suggests that CONSISTENCY is harder than *SHIFT and confirms what is normally assumed in the literature.

***TRIVIAL** is a constraint which forbids under-informative interpretations (see Stalnaker's first principle of rational conversation). Recall the original workshop example discussed in chapter 1 in which *TRIVIAL conflicts with *SHIFT.

the workshop You are attending a workshop. In front of you lies the list of names of all participants, around you are sitting the participants in flesh and blood. Consider the following dialogue, question, and assertion, uttered in such a situation:

(216) A: Who _{n} is that man?

B: That man is Ken Parker.

A: Who _{m} is Nathan Never?

B: Nathan Never is the one over there.

(217) Who _{n} is who _{m} ?

(218) I don't know who _{n} is who _{m} .

Again if we follow *SHIFT, then n and m should be assigned the same value. On the other hand, if *SHIFT is satisfied, *TRIVIAL which forbids trivial interpretations, would be violated. Since, intuitively, we shift conceptualization while interpreting these sentences, the example suggests that *TRIVIAL ranks higher than *SHIFT.

RELEVANCE expresses preference for relevant interpretations (see Grice's maxim of relation and Horn's I-principle). A formal characterization of the notion of relevance has been recently attempted by a number of authors, for instance Roberts (1996b), Groenendijk (1999), and in particular van Rooy (2000a), who also discusses examples showing how relevance can influence the interpretation of belief attributions.¹⁵ I will not discuss these formalizations though, because the intuitive notion of relevance which Grice had in mind, suffices to make my point here. Consider the following slight modification of the Ann-Bea example we discussed in chapter 2.

Ann and Bea In front of Ralph stand two women. For some reason we don't need to investigate, Ralph believes that the woman on the left, who is smiling, is Bea and the woman on the right, who is frowning, is Ann. As a matter of fact, exactly the opposite is the case. Bea is frowning on the right and Ann is smiling on the left. In the picture w_1 is the world of evaluation and w_2 is the only possibility in Ralph's belief state.

$$\begin{array}{ccc} w_1 & \mapsto & \begin{array}{cc} (\text{☺}) & (\text{☹}) \\ [ann] & [bea] \end{array} \\ \\ w_2 & \mapsto & \begin{array}{cc} (\text{☺}) & (\text{☹}) \\ [bea] & [ann] \end{array} \end{array}$$

There are two possible conceptual covers in such a situation, namely the set $A = \{\lambda w[\text{left}]_w, \lambda w[\text{right}]_w\}$, which cross-identifies the women which stand in the same perceptual relation to Ralph and the set $B = \{\lambda w[\text{Ann}]_w, \lambda w[\text{Bea}]_w\}$, which cross-identifies the women by their name. Suppose all of a sudden Ralph starts chasing the woman on the left to bring her to a mental institution. I ask you: 'Why is Ralph chasing Ann?'. You answer:

(219) Ralph believes that Ann is insane.

There are three possible ways of interpreting this sentences in the described situation: (a) an interpretation *de re* under cover A, in which Ann is identified as the woman on the left; (b) an interpretation *de re* under cover B, in which Ann is identified as Ann; (c) the *de dicto* interpretation.

All three interpretations are consistent with the background. Interpretation (a) seems to involve a violation of *SHIFT, and probably STRENGTH. Indeed, my question, which explicitly uses 'Ann' to identify the relevant woman, suggests cover B as the prominent one. (b) and (c) do not involve such violation(s). Still, intuitively, we prefer interpretation (a) for (219) in such situation. I suggest

¹⁵The Ann-Bea case I discuss here has been inspired by van Rooy's example of the English gentleman.

that the reason is that only under such an interpretation the sentence would be relevant. Indeed, whether the belief attribution (219) is contributing to explain for us Ralph's behaviour depends on how Ann is identified in Ralph's belief state. Whether or not Ralph believed that Ann – who is, according to him, the woman on the right – is insane does not help explaining why he is chasing the woman on the left, whereas the fact that he believes that the woman on the left is insane does contribute to an explanation. Thus, only under interpretation (a) the belief attribution constitutes a proper answer to my question and hence is relevant. This is why the addressee adopts a cover containing the concept 'the woman on the left' in such a situation, although this involves a violation of *SHIFT. By assuming that RELEVANCE is harder than *SHIFT we can account for this intuition.

Synopsis

We have discussed the following interpretation constraints which seem to play a role in the operation of cover selection:

ANCHOR Interpretation should be anchored to the context.

CONVERSATIONAL MAXIMS

CONSISTENCY avoid inconsistent interpretations;

***TRIVIAL** avoid trivial interpretations;

RELEVANCE prefer relevant interpretations.

***SHIFT** Do not shift conceptualization.

***ACCOMMODATION** Do not accommodate.

STRENGTH Stronger covers are preferred.

The following is a possible ranking consistent with the phenomena discussed above:

ANCHOR, C. MAXIMS > *SHIFT, *ACCOMMODATION > STRENGTH

If we order the principles we discussed in this way,¹⁶ we are able to explain the process of saturating *CC*-indices in many of the cases we have discussed in the thesis. As a further illustration I will consider the case of Susan's mother, discussed in chapter 2. The interpretation procedure in this example involves a number of contrasting constraints and clearly illustrates the nature of an optimality theoretic explanation.

¹⁶By the assumption of other ordered constraints we might have reached exactly the same result. This is not really important here, what is relevant is that constraints play a role here which can be overruled by higher ranking ones.

Susan's mother For ease of reference, I rewrite the relevant situation described by van Fraassen (1979):

Susan's mother is a successful artist. Susan goes to college, where she discusses with the registrar the impact of the raise in tuition on her personal finances. She reports to her mother 'He said that I should ask for a larger allowance from home'. Susan's mother exclaims:

(220) He must think I am rich.

Susan, looking puzzled, says 'I don't think he has any idea who you are'.¹⁷

van Fraassen analyzes the example as follows:

The information the mother intends to convey is that the registrar believes that Susan's mother is rich, while Susan misunderstands her as saying that the registrar thinks that such and such successful artist is rich. The misunderstanding disappears if the mother gives information about herself, that is, about what she had in mind. She relied, it seems, on the auxiliary assertion 'I am your mother'.¹⁸

A number of conflicting constraints play a role in the interpretation of (220). On the one hand, we have *SHIFT, *ACCOMMODATION and STRENGTH which forbid the selection of a cover containing the concept 'Susan's mother', because it is not prominent and clearly weaker than other salient ones with respect to the registrar's information state. On the other hand, we have CONSISTENCY which, if fulfilled, forces precisely such a non-salient and weak interpretation. Susan accepts the sentence after the clarification of the mother, and this confirms our hypothesis that CONSISTENCY can overrule *SHIFT, *ACCOMMODATION and STRENGTH.

Suppose now that Susan's mother utters the following sentence in the same situation:

(221) He must think I am your mother.

Only under a conceptualization containing the concept 'Susan's mother', the belief attribution in (221) is consistent with the common ground, but, under such a cover, the sentence is clearly trivial. Thus, the only way to satisfy CONSISTENCY here would involve a violation of *TRIVIAL. This explains why the sentence is pragmatically unacceptable in such a situation. Still, it seems that the inconsistent reading is preferred over the trivial one in this case and the present analysis can explain this fact as follows. While the inconsistent reading

¹⁷van Fraassen (1979), pp. 371-372 and chapter 2.

¹⁸van Fraassen (1979), p. 372.

just violates CONSISTENCY, the trivial reading violates *TRIVIAL, but also *SHIFT, *ACCOMMODATION and STRENGTH, and these violations of lower constraints become crucial in this case.

The following two diagrams summarize our OT-analysis of the case of Susan's mother. I use (*) to indicate that the interpretation violates the corresponding constraint, and !(*) to indicate a crucial violation. Optimal interpretations are those which do not involve any crucial violation.

(220) He must think I am rich.

- (a) *de re* interpretation under the prominent cover containing the concept 'such and such successful artist': $\lambda x \Box R(x)(a)$;
- (b) *de re* interpretation under a cover containing the concept 'Susan's mother': $\lambda x \Box R(x)(m)$.

(220)	*TRIV, CONS	*SHIFT, *ACC	STRENGTH
(a)	!(*)		
(b)		(*) (*)	(*)

(221) He must think I am your mother.

- (a) *de re* interpretation under the prominent cover containing the concept 'such and such successful artist': $\lambda x \Box M(x)(a)$
- (b) *de re* interpretation under a cover containing the concept 'Susan's mother': $\lambda x \Box M(x)(m)$

(221)	*TRIV, CONS	*SHIFT, *ACC	STRENGTH
(a)	(*)		
(b)	(*)	!(*), !(*)	(*)

We will return to this example later on.

Bi-Dimensional Optimality Theory

The OT analysis discussed so far enables an explanation of the process of saturating *CC*-indices in many of the cases we have discussed in the thesis, but not all. As an illustration consider the following example.

the bald president Consider the following situation. Naming is the prominent cover and the addressee holds as common ground that: (i) Putin is the actual president of Russia; (ii) Ralph believes that Jeltsin is the actual president of Russia; (iii) Ralph would not assent to the sentence: 'Putin is bald'. Consider now the following sentence uttered in such a situation:

(222) Ralph believes that Putin is bald.

Let A be naming and B be a cover containing the concept ‘the actual president of Russia’. The sentence has three possible interpretations in such circumstances: (a) the *de dicto* reading; (b) the *de re* reading under A; and (c) the *de re* reading under B. According to the interpretation theory I have discussed so far, interpretation (c) would be optimal given the situation. Indeed, although such an interpretation violates *SHIFT, *ACCOMMODATION (and probably STRENGTH), the other two alternative interpretations, which satisfy these constraints, crucially violate CONSISTENCY, because of clause (iii) above.

(222)	CONS	*SHIFT,	*ACC	STRENGTH
(a) <i>de dicto</i>	!(*)			
(b) <i>de re</i> -naming	!(*)			
(c) <i>de re</i> -descript.		(*)	(*)	(*)

As illustrated by the constraint table, interpretation (c) is the predicted optimal interpretation. On such interpretation, the sentence says that (Putin is the actual president of Russia and) Ralph would assent to the sentence ‘The actual president of Russia is bald’. Since Ralph believes that Jeltsin is the actual president of Russia, (c) also entails the *de dicto* reading of the sentence: ‘Ralph believes that Jeltsin is bald’. This prediction is clearly counter-intuitive. An intuitive explanation of why reading (c) is not preferred in such a situation is that a speaker expressing such a content by means of such a sentence would not be cooperative. Indeed, in the described situation, the same content could have been conveyed in a much more efficient way by uttering the following sentence:

(223) Ralph believes that the president of Russia is bald.

The *de dicto* reading of this alternative formulation and reading (c) of (222) convey the same information in the described situation in which the information that Putin is the actual president of Russia is part of the common ground. But the former interpretation does not involve any shift of cover or accommodation. For this reason, (223) is more efficient than (222), and, therefore, the speaker, if cooperative, should have chosen it. This is Grice’s principle of cooperation. A speaker has a responsibility of what the audience will make of her sentences. In cooperative exchanges, she goes through the interpretation herself and makes sure that the intended content is as easy to obtain as possible. A cooperative speaker would never have uttered (222) to convey the information that Ralph would assent to the sentence ‘Jeltsin is bald’. Therefore an interpretation of (222) which conveys such information cannot be optimal in such a situation. Note, however, that such an explanation cannot be formulated in the OT interpretation theory we have considered so far, in which inputs are given by single sentences and no reference is made to alternative sentences that the speaker might have used. In order to account for these cases, we need a more complex analysis, where the optimal solution is searched on two dimensions, rather than one: the dimension of the addressee and the one of the speaker, and in which the two

optimization procedures of the addressee and of the speaker can refer to each other and crucially constrain each other. Such an analysis is the bi-directional Optimality Theory of Reinhard Blutner (see Blutner (1999) and Blutner and Jäger (1999)). In the next section, I follow Dekker and van Rooy (1999) and define bi-directional OT interpretation as an ‘interpretation game’. The use of game-theoretical concepts allows a perspicuous formulation of Blutner’s central notions of strong and weak optimality. Furthermore, a game-theoretic formulation of the process of interpretation will be useful in order to account for the interplay between the addressee and the speaker with their particular interests and goals in the interpretation of context dependent natural language expressions.

4.3.2 Interpretations as Games

In an interesting recent article,¹⁹ Dekker and van Rooy (D&vR) propose to apply concepts that have been studied in the field of Game Theory²⁰ to investigate a series of phenomena in the semantic/pragmatic interface. They rewrite OT interpretation theory in terms of game-theoretic notions where optimality itself is viewed as a solution concept for a game.

The central notion introduced by D&vR is that of an interpretation game. Interpretation games are defined in terms of ‘strategic games’.

A *strategic game* G is a triple

$$G = (N, (A_i)_{i \in N}, (\succ_i)_{i \in n})$$

where N is a set of players, $(A_i)_{i \in N}$ maps each player to a non-empty set of alternative actions A_i , and \succ_i is a preference relation for player i over the product $A = A_1 \times \dots \times A_n$ of possible actions of all players. An element a of such a product A is called an *action profile*.

An *interpretation game* I is a strategic game involving two players, the Speaker and the Hearer, $N = \{S, H\}$. The set of alternative actions for the speaker consists of a set $A_S = \{F1, F2, \dots\}$ of possible *forms*, the set of alternative actions for the hearer consists of a set $A_H = \{C1, C2, \dots\}$ of possible *contents*. S chooses a suitable form $F \in A_S$ for a content $C \in A_H$ to be communicated. H chooses a suitable interpretation $C \in A_H$ for a signaled representation $F \in A_S$. Optimality theoretic preferences are used in combination with particular goal-directed preferences to define the preference relations of the speaker \succ_S and of the hearer \succ_H . The relations \succ_S and \succ_H should be interpreted as strict preferences and hence are taken to be transitive, anti-reflexive and anti-symmetric.

The interpretation games I will consider are crucially played in a specific context. I will therefore identify forms with utterances and interpretations with

¹⁹Dekker and van Rooy (1999).

²⁰See Osborne and Rubinstein (1994) for an introduction to the main game-theoretical concepts.

actions on the specific information state which constitutes the common-ground in the circumstances of these utterances (see Blutner (1999) who makes the same assumption). The fulfillment of the interpretation constraints discussed in the previous section and general principles of cooperation will determine the preference relation of the hearer \succ_H . General principles of generation, cooperativity and the particular goals of the speaker will interplay in the determination of \succ_S . These preference relations are highly context dependent. In the present analysis, they are sensitive to three specific aspects of the context: (a) which covers are salient; (b) which information is presupposed by the speaker and by the addressee; (c) the specific intentions of the speaker. The first two factors are relevant in that they determine whether or not a profile satisfies the interpretation constraints discussed above, and hence influence the preference relation of the addressee and of the speaker, if cooperative; the third factor helps in determining which content is intended by the speaker who has authority on how her utterance should be interpreted, and hence influences the preference relation of the speaker.²¹

One of the tasks of game theory is to develop solution concepts which allow one to make predictions about the outcome of a game and about how the players will interact. In D&vR, optimality is viewed as a solution concept of an interpretation game. Optimal solutions are no longer optimal interpretations of a given expression, but optimal profiles consisting of an utterance and an interpretation. D&vR discuss two notions of optimality: the notion of *Nash-optimality*, and *BJ-optimality*. The first notion is nothing else than the well-known solution concept of Nash equilibrium, which is shown to be the game-theoretic equivalent of Blutner's (1999) notion of strong optimality. BJ-optimality is the game-theoretic counterpart of Blutner's notion of weak optimality. I will present D&vR's definitions of these two notions. I will then show how these two solution concepts can be used to account for our intuitions about the bald president example discussed above.

Nash- and BJ-optimality

In Blutner's (1999) bi-directional OT, a mechanism compares different possible interpretations C for the same syntactic expression F and another mechanism compares different possible syntactic formulations F for the same content C. A form-content pair (F,C) is then *strongly optimal* just in case C is an optimal interpretation for F according to the first mechanism and F is an optimal form

²¹Eventually the specific intentions of the speaker might also influence the preference relation of the addressee, if cooperative. This might be the sense of cooperativity from the addressee point of view. Being cooperative means to minimize effort for the other participant. S is cooperative, if she chooses the form which can be interpreted in the most straightforward way by H. H cannot do much to help S in choosing the right words, but she can be cooperative by selecting the intended interpretation. Obviously only correctly informed agents can be cooperative. If S and H fail to share a common-ground or if H fails to know the intentions of S, misunderstanding can arise and communication can break down.

for C according to the second mechanism. D&vR have shown that this notion of strong optimality can be perspicuously formalized by means of the classical solution concept of a Nash equilibrium. Blutner's strong optimal solutions are identified with Nash equilibria in an interpretation game.

Given an action profile $a \in A$ and an action $a_i \in A_i$, let $a[i : a_i]$ denote the profile which is like a , but with player i taking action a_i .

4.3.4. DEFINITION. [Nash-optimality] Let $I = (N, (A_i)_{i \in N}, (\succ_i)_{i \in N})$ be an interpretation game. An action profile a is *Nash-optimal* in I , $\text{NASH}_I(a)$ iff

$$\forall i \in N : \forall a_i \in A_i : \neg(a[i : a_i] \succ_i a)$$

Intuitively, in a Nash equilibrium, every player acts optimally given the other players' actions, that is, every player's action is the best response to the choices of the other players.

As an illustration consider the interpretation game depicted by means of the following matrix:

	<i>C1</i>	<i>C2</i>
<i>F1</i>	(2, 3)	(4, 5)
<i>F2</i>	(3, 2)	(1, 1)

In such matrices, the Speaker chooses the row and the Hearer the column to be played and preference relations are formulated in terms of payoff functions,²² where the payoff pair (x, y) expresses that the speaker gets payoff x and the hearer gets payoff y . In the game depicted by this specific matrix, S prefers causing *C1* by performing *F2*, whereas *F1* is the preferred way of causing *C2*. H prefers performing action *C2* as a response to *F1*, and *C1* as a response to *F2*. The game has two Nash equilibria, namely the profiles $(F2, C1)$ and $(F1, C2)$.

In the definition of a Nash equilibrium the only strict preferences \succ_i which really count are those between two profiles a and b if their only difference lies in the choice of $i \in \{S, H\}$, i.e. if $a = b[i : b_i]$ for some b_i . For this reason D&vR propose to represent Nash equilibria in interpretation games²³ by drawing arrows between two profiles on the same row or in the same column, with the following meaning: \leftarrow means 'H strictly prefers the left profile', \rightarrow means 'H strictly prefers the right profile', \downarrow 'S strictly prefers the bottom profile', and \uparrow 'S strictly prefers the top profile'. The game above is then represented by the following table in which the Nash equilibria are immediately visualized by \circ :

	<i>C1</i>	<i>C2</i>
<i>F1</i>	\rightarrow	\circ
<i>F2</i>	\downarrow	\uparrow
	\circ	\leftarrow

²²Preference relations can be expressed in terms of payoff functions $(u_i)_{i \in N}$, where $u_i : A \rightarrow R$ is the payoff function of player i . Action profiles with higher payoff are preferred.

²³Interpretation games crucially involve only two players.

If no arrow is leaving from a profile a , then a is a Nash equilibrium. This means that a profile (F, C) is Nash-optimal in I iff for all contents $CN \in A_H$ and forms $FN \in A_S$ in I :

- (i) $(F, CN) \not\prec_H (F, C)$
- (ii) $(FN, C) \not\prec_S (F, C)$

By means of the notion of Nash-optimality, we can characterize anomalous interpretations. Intuitively, a pair (F, C) is anomalous with respect to I iff it is *not* Nash-optimal in I , and this is the case iff either C is not an optimal interpretation for F in I (clause (i) is not satisfied) or, if C is an optimal interpretation, then C could have been expressed more efficiently by an alternative form (clause (ii) is not satisfied).

The strong version of optimality characterized by the notion of a Nash equilibrium is useful to explain many standard cases, but it has been shown not to be always satisfactory. In his (1999) article, Blutner illustrates this by means of the following example inspired by Horn:

(224) Black Bart killed the sheriff.

(225) Black Bart caused the sheriff to die.

The lexical causative *kill* tends to be restricted to stereotypical causative situations (e.g. Black Bart shot the sheriff), and the marked construction in (225) tends to refer to more marked situations (e.g. Black Bart caused the sheriff's gun to backfire by stuffing it with cotton). The general tendency illustrated by this example seems to be that 'unmarked forms tend to be used for unmarked situations and marked forms for marked situations' (Horn (1984), p. 26). This tendency has been called by Horn *the division of pragmatic labour*.

This case can be formalized by means of the following interpretation game:

	$C1$	$C2$
$F1$	\rightarrow	
	\downarrow	\downarrow
$F2$	\rightarrow	\circ

where $F1$ and $F2$ stand for the marked and the unmarked forms respectively and $C1$ and $C2$ stand for the the marked and the unmarked situation respectively. By the notion of Nash-optimality we can account for the fact that (224) picks up stereotypical situations. Unmarked forms ($F2$) are preferred over marked forms ($F1$), and stereotypical situations ($C2$) are easier to understand than atypical situations ($C1$). The profile unmarked form-unmarked situation ($F2, C2$) is Nash in such a game. But Nash-optimality is not sufficient to explain why (225) obtains the unusual interpretation. Indeed, no interpretation is selected for the marked

form $F1$. The profile marked form-marked content is intuitively chosen because (i) the alternative unmarked form does not get the marked interpretation and (ii) we prefer to use the unmarked form to express the unmarked situation. Now, by means of the notion of optimality defined in terms of a Nash equilibrium we cannot capture this kind of reasoning. A profile is Nash-optimal iff it is optimal for the Speaker and optimal for Hearer and these two checks for optimality are independent of each other. The search for the optimal choice for one player is not influenced by the preference relation of the other player. In order to account for H's reasoning in this case, we need a notion in which the two optimization procedures of the hearer and of the speaker can refer to each other and constrain each other. Such a notion is the notion of *weak optimality* introduced in Blutner (1998). D&vR's notion of BJ-optimality is the perspicuous game-theoretical formulation of such notion.

BJ-optimality is defined as follows (see Dekker and van Rooy (1999) for further discussion):

4.3.5. DEFINITION. [BJ-Optimality] Let $I = (N, (A_i)_{i \in N}, (\succ_i)_{i \in N})$ be an interpretation game. Then the set BJ_I of BJ-optimal solutions in I is defined as follows:

$$BJ_I = NASH_{I_n}$$

where I_n is the fixed point, i.e. $I_{n+1} = I_n$, of the sequence of games I_0, \dots, I_m, \dots constructed as follows:

- (i) $I_0 = I$
- (ii) $I_{n+1} = (N, (A)_{i \in N}, (\succ_{i_{n+1}})_{i \in N})$ with
 - (a) $\succ_{s_{n+1}} = \succ_{s_n} \setminus \{(y, z) \mid \exists x \in NASH_{I_n} : x \succ_{H_n} y\}$;
 - (b) $\succ_{H_{n+1}} = \succ_{H_n} \setminus \{(y, z) \mid \exists x \in NASH_{I_n} : x \succ_{S_n} y\}$.

In the construction of I_{n+1} you eliminate preferences for blocked profiles y . A profile y is blocked in I_{n+1} , if there was a Nash-optimal profile x which was preferred to y in I_n . If $I_{n+1} = I_n$, then the Nash equilibria of I_n are the BJ-optimal solutions in I_0 . That is, if an action profile a is a Nash-optimal solution in the fixed point game of the sequence generated from a game I , then a is BJ-optimal in I .

The intuitive idea of this construction is that Nash-optimal profiles block less preferred ones and preferences for blocked profiles are overruled. As an illustration, let's go back to the game I determined by Horn's sheriff example. The sequence generated from such game consists of the two games represented in the following matrices where blocked profiles are indicated by \perp :

		<i>C1</i>		<i>C2</i>			<i>C1</i>		<i>C2</i>																					
I_0 :	<i>F1</i>	<table> <tr><td></td><td>\rightarrow</td><td>\perp</td></tr> <tr><td>\downarrow</td><td></td><td>\downarrow</td></tr> <tr><td>\perp</td><td>\rightarrow</td><td>\circ</td></tr> </table>					\rightarrow	\perp	\downarrow		\downarrow	\perp	\rightarrow	\circ		I_1 :	<i>F1</i>	<table> <tr><td>\circ</td><td></td><td>\perp</td></tr> <tr><td></td><td></td><td>\downarrow</td></tr> <tr><td>\perp</td><td>\rightarrow</td><td>\circ</td></tr> </table>				\circ		\perp			\downarrow	\perp	\rightarrow	\circ
	\rightarrow	\perp																												
\downarrow		\downarrow																												
\perp	\rightarrow	\circ																												
\circ		\perp																												
		\downarrow																												
\perp	\rightarrow	\circ																												
	<i>F2</i>					<i>F2</i>																								

I_0 is I , and I_1 is obtained from I_0 by eliminating preferences for the two blocked profiles $(F2, C1)$ and $(F1, C2)$. I_1 has two Nash-optimal solutions: $(F1, C1)$ and $(F2, C2)$. Since no preference can be eliminated in the next step of our construction (i.e. $I_1 = I_2$), these two profiles are the two BJ-optimal solutions of I .

	<i>C1</i>	<i>C2</i>
<i>F1</i>	BJ \rightarrow	\perp
	\downarrow	\downarrow
<i>F2</i>	$\perp \rightarrow$	\circ

The profiles $(F1, C2)$ and $(F2, C1)$ are not BJ-optimal in I , because overruled by the Nash-optimal $(F2, C2)$. On the other hand, $(F1, C1)$ is BJ-optimal. Although $(F1, C2) \succ_H (F1, C1)$ and $(F2, C1) \succ_S (F1, C1)$, these preference do not count because $(F1, C2)$ and $(F2, C1)$ are blocked by the Nash-optimal $(F2, C2)$, since $(F2, C2) \succ_S (F1, C2)$ and $(F2, C2) \succ_H (F2, C1)$. In the notion of BJ-optimality, a player's perspective on optimization is crucially constrained by the other player's perspective and *vice versa*. We thus capture H's intuitive reasoning in the sheriff case. The profile marked form-marked content $(F1, C1)$ is intuitively chosen because (i) the alternative unmarked form $F2$ does not get the marked interpretation $C1$ ($(F2, C2) \succ_H (F2, C1)$) and (ii) we prefer to use the unmarked form $F2$ to express the unmarked situation $C2$ ($(F2, C2) \succ_S (F1, C2)$). The hearer chooses the marked $C1$ rather than the unmarked $C2$ as interpretation for $F1$, because she can reason as follows: if the speaker had wanted to communicate $C2$ he would have chosen the Nash-optimal $F2$.

Now let us see how the bald president case discussed above can be accounted for by means of these notions.

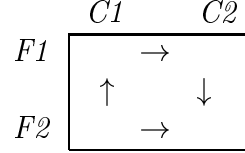
the bald president For ease of reference, I restate the situation. Naming is the prominent cover. The common ground contains the following information: (i) Putin is the actual president of Russia; (ii) Ralph believes that Jeltsin is the actual president of Russia; (iii) Ralph would not assent to the sentence: 'Putin is bald'. In such context, the following sentence is uttered by the speaker S:

(222) Ralph believes that Putin is bald.

Intuitively, a rational addressee H can do two things in such a situation: either refute to perform any action or consider revising her state with the information that Ralph does believe *de dicto* that Putin is bald. In any case, H does not

update with the information that Ralph would assent to the sentence: ‘Jeltsin is bald’. This last action was predicted as optimal by the one-dimensional OT interpretation theory I introduced in the previous section. Let’s see whether the two-dimensional theory I have just described does any better here.

I propose to characterize such a situation by means of the following interpretation game:



F1 and *F2* are the utterances of the sentences:

(222) Ralph believes that Putin is bald.

(223) Ralph believes that the actual president of Russia is bald.

For ease of reference, I will denote the first action by ‘*Putin*’ and the second action by ‘*the president*’.

F1 = ‘*Putin*’

F2 = ‘*the president*’

Let us see now how *C1* and *C2* are characterized. Let *A* be naming and *B* be a cover containing the concept ‘the actual president of Russia’. Assume *A* and *B* are the only two covers available in our situation. Each of the two sentences above has then three possible interpretations. Let \wp be a conceptual perspective such that $\wp(n) = A$ and $\wp(m) = B$.

(222) Ralph believes that Putin is bald.

- a. *de dicto*: $\Box B(p)$
- b. *de re* under *A*: $\exists x_n[x_n = p \wedge \Box B(x_n)]$
- c. *de re* under *B*: $\exists x_m[x_m = p \wedge \Box B(x_m)]$

(223) Ralph believes that the actual president of Russia is bald.

- d. *de dicto*: $\Box B(r)$
- e. *de re* under *A*: $\exists x_n[x_n = r \wedge \Box B(x_n)]$
- f. *de re* under *B*: $\exists x_m[x_m = r \wedge \Box B(x_m)]$

Given our characterization of the situation, these six possible interpretations collapse in only two different possible actions on the relevant common ground. Let σ stand for the common ground in the described situation and let \wp be as above, we then obtain the following equivalences:²⁴

²⁴Recall that σ supports the information that Putin is the actual president of Russia, $r = p$.

$$(\alpha) \ UP_{\sigma}^{\varphi}(a) = UP_{\sigma}^{\varphi}(b) = UP_{\sigma}^{\varphi}(e)$$

$$(\beta) \ UP_{\sigma}^{\varphi}(c) = UP_{\sigma}^{\varphi}(d) = UP_{\sigma}^{\varphi}(f)$$

The updates of σ with the six interpretations above collapse in only two possible updates. There are only two possible alternative actions for H on the relevant state: the action in (α) which consists in eliminating those possibilities in σ in which it is true that Ralph believes *de dicto* that Putin is bald; the action in (β) which consists in eliminating those possibilities in σ in which it is true that Ralph believes *de dicto* that the actual president of Russia is bald. For ease of reference, I will denote the first action by $UP(put)$ and the second action by $UP(pres)$. I will identify $C1$ and $C2$ with these two possible updates:

$$C1 = UP(put)$$

$$C2 = UP(pres)$$

Let us turn now to the preference relations \succ_H and \succ_S . H's preferences are obtained by the following two OT-constraint tables for the two relevant utterances and contents:

F1	CONS	*SHIFT,	*ACC	STRENGTH
C1	!(*)			
C2		(*)	(*)	(*)

F2	CONS	*SHIFT,	*ACC	STRENGTH
C1	!(*)			
C2				

$C1$ crucially violates CONSISTENCY in both diagrams because such an action on the common ground leads to the absurd state in any case. Therefore, in our game, H strictly prefers a profile $(FN, C2)$ over $(FN, C1)$. Consistent interpretations are preferred over inconsistent interpretations.

As for the speaker's preferences, I assume in the matrix that S strictly prefers causing $C1$ by performing $F1$ and causing $C2$ by performing $F2$. Indeed, an utterance of the sentence: 'Ralph believes that Putin is bald' (i.e. $F1$) is the most cooperative way of conveying the information that Ralph would assent to the sentence: 'Putin is bald' (i.e. the information brought about by $C1$) and an utterance of the sentence: 'Ralph believes that the actual president of Russia is bald' (i.e. $F2$) is the most cooperative way of conveying the information that Ralph would assent to the sentence: 'The actual president of Russia is bald' (i.e. the information brought about by $C2$). Therefore, if we assume that S is cooperative, we obtain the following preference relation \succ_S :

$$('Putin', UP(put)) \succ_S ('the\ president', UP(put))$$

$$('the\ president', UP(pres)) \succ_S ('Putin', UP(pres))$$

Cooperative formulations are preferred over non-cooperative formulations. While the first preference does not have any effect on the following discussion, the second preference is crucial to the outcome of the game.²⁵

Our game has one Nash equilibrium, namely the profile $('the\ president', UP(pres))$:

	$UP(put)$	$UP(pres)$
$'Putin'$	\rightarrow	
	\uparrow	\downarrow
$'the\ president'$	\rightarrow	\circ

Nash-optimality selects $UP(pres)$ as optimal interpretation for $'the\ president'$ because consistent interpretations are preferred over inconsistent interpretations and efficient formulations are preferred over non-efficient formulations. But Nash-optimality does not select any interpretation for $'Putin'$. Its interpretation is left open because (i) profile $('Putin', UP(put))$ is anomalous since $UP(put)$ is not an optimal interpretation (it leads to inconsistency) and (ii) $('Putin', UP(pres))$ is anomalous since although $UP(pres)$ is an optimal interpretation, it could have been expressed more efficiently by an alternative form. In order to account for the fact that the inconsistent interpretation of (222) under the prominent cover is preferred by the hearer over an interpretation under the problematic conceptualization we need the weaker notion of BJ-optimality. Intuitively the hearer chooses action $UP(put)$, which would lead her to the absurd state, rather than $UP(pres)$ as a response to $'Putin'$ because she can reason as follows: If the Speaker had wanted to convey the consistent interpretation $UP(pres)$, then S should have chosen the more efficient formulation $'the\ president'$. But S chose $'Putin'$. Thus S must have meant to convey $UP(put)$. This is precisely the kind of reasoning captured by the notion of BJ-optimality. Indeed, profile $('Putin', UP(put))$ is BJ-optimal in our game. Whereas $('Putin', UP(pres))$ is not, because overruled by the Nash-optimal $('the\ president', UP(pres))$.

	$UP(put)$	$UP(pres)$
$'Putin'$	BJ \rightarrow	\perp
	\uparrow	\downarrow
$'the\ president'$	$\perp \rightarrow$	\circ

²⁵This preference is further justified by the constraint table above. Indeed, $C2$ violates *SHIFT, *ACC and STRENGTH in the first diagram, but not in the second. This is due to the fact that according to our semantics $F1$ can produce in σ the update effect $C2$ only by means of an update with reading (c), which, as we have seen, involves such violations. Whereas $F2$ can convey such information by means of reading (d), which does not involve any constraint violation. Therefore, the speaker, who is assumed to be cooperative, crucially prefers profile $(F2, C2)$ over $(F1, C2)$.

We can now explain the addressee's behaviour in our presidential example. In order to interpret an utterance of 'Ralph believes that Putin is bald', the addressee does not adopt a cover containing the concept 'the actual president of Russia' (profile ('*Putin*', *UP(pres)*) is blocked), but she rather assumes the prominent conceptualization (profile ('*Putin*', *UP(put)*) is BJ-optimal). The latter action leads her to the absurd state. She can protest or she can decide to start a process of revision of her information.

How is the bald president situation different from the case of Susan's mother that we have considered above? As in the presidential example, in the case of Susan's mother we have a sentence whose only interpretation which does not contradict the common ground involves a violation of *SHIFT, *ACCOMMODATION and STRENGTH. In the presidential example, such an interpretation was intuitively unacceptable and we could account for it. Instead, in the case of Susan's mother, such an interpretation is intuitively acceptable. We have to show how it is possible that the violation of exactly the same interpretation principles can be tolerated on one occasion and not on the other. Although the two cases have a similar structure, there is a crucial difference. In the latter case, we can find a series of reasons justifying the Speaker's chosen utterance, and hence, although the chosen formulation of the intended content is clearly uncooperative, it is hard to find an alternative formulation which is *strictly* preferred by S over the one actually used. This, I suggest, explains our different intuitions about the addressee reactions in the two cases. Let us have a closer look.

Susan's mother again For ease of reference, I restate the situation:

Susan's mother is a successful artist. Susan goes to college, where she discusses with the registrar the impact of the raise in tuition on her personal finances. She reports to her mother 'He said that I should ask for a larger allowance from home'. Susan's mother exclaims: 'He must think I am rich.' Susan, looking puzzled, says 'I don't think he has any idea who you are'.²⁶

van Fraassen analyzes the example as follows:

The information the mother intends to convey is that the registrar believes that Susan's mother is rich, while Susan misunderstands her as saying that the registrar thinks that such and such successful artist is rich. The misunderstanding disappears if the mother gives information about herself, that is, about what she had in mind. She relied, it seems, on the auxiliary assertion 'I am your mother'.²⁷

²⁶ van Fraassen (1979), pp. 371-372 and chapter 2.

²⁷ van Fraassen (1979), p. 372.

If you follow the line of reasoning I used above, you may think that in order to make life easy for Susan, the mother should have used (226) instead of (227) in order to convey the intended information:

(226) He must think that your mother is rich.

(227) He must think I am rich.

The first formulation is naturally interpreted *de dicto*, and, on such a reading, it does not involve any violation of interpretation constraints. Thus, it is a more cooperative formulation than (227), since the intended interpretation of the latter involves a violation of *SHIFT, *ACCOMMODATION and STRENGTH in the described situation (see the OT-table relative to this example in the previous section). Still Susan's mother chooses to utter (227) and her choice is not unmotivated. First of all, Susan's mother is personally involved and probably upset because of the raise in tuition. It is *her* who has to pay more money now and the use of the personal pronoun 'I' rather than the neutral description 'your mother' is a more effective way to express her personal commitment and feelings about the situation. Furthermore, we might also assume that the following generation principle plays a role here, Zeevat (1999b) calls it **ParsePerson**, which expresses a preference for the use of the personal pronouns 'I', 'you' instead of descriptions in order to refer to the participants to the conversation. Principle **ParsePerson** might be related to a more general principle which expresses a preference for shorter utterances which require less effort to be generated. An utterance of (226) would violate such a principle whereas an utterance of (227) would not. From these considerations, it seems fair to conclude that in contrast to the previous presidential case where cooperative formulations were strictly preferred by the speaker, here, given the particular circumstances, none of the two formulations is strictly preferred over the other from Susan's mother perspective: the use of the description 'your mother' is more cooperative, but the use of the personal pronoun 'I' (i) is more effective for her in this particular situation in order to express her feelings and (ii) is shorter and thus possibly preferred because requires less effort to be generated.

This informal discussion suggests the following formalization of Susan's interpretation problem in such a situation:

	$UP(art)$		$UP(moth)$
'I'	\perp	\rightarrow	\circ
	\uparrow		
'your mother'	\perp	\rightarrow	\circ

where

$UP(art)$ is the result of updating the common ground with *de re* reading of (227) under the prominent cover containing the concept 'such and such a famous artist'.

$UP(moth)$ is the result of updating the common ground with *de re* reading of (227) under a cover containing the concept ‘Susan’s mother’.

Note that $UP(moth)$ is equivalent to an update of the common ground with the *de dicto* reading of (226).

$UP(moth)$ is preferred over $UP(art)$ by Susan because the latter would lead to inconsistency. The use of the description ‘*your mother*’, although it is more cooperative, is not strictly preferred over the use of the personal pronoun ‘*I*’ by Susan’s mother because of the reasons discussed above.

Susan’s mother utterance and her intended interpretation are Nash-optimal in such a game, as well as the profile containing the alternative cooperative formulation. The other two profiles are not even BJ-optimal.

4.4 Conclusion

In the first part of this chapter, I have looked more specifically at a number of formal properties of conceptual covers and I have compared them with alternative notions of cross-world identification. In the second part, I have attempted a first description of the pragmatics of conceptual covers. The proposed analysis is certainly not conclusive. A number of open questions remain in connection to the OT analysis discussed in the first part, for instance, concerning the choice of the constraints and their ranking. As for the game theoretical part, the field is new and most of the theoretical questions are still unsettled. For instance, already the characterization of the basic ingredients of an interpretation game is open to discussion. My identification of the set of possible actions for the addressee A_H with the set of possible update effects on the current common ground is not without consequences and is in need of further justification. Another open question is how generation constraints, interpretation principles and particular goals combine in the determination of the preference relations. All these issues are challenging and ask for further investigation, which, however, must be left to another occasion.