Introduction to MSDRT

Hans Kamp and Brendan Learnihan-Sylvester

November 13, 2022

Contents

Chapter 1

A very brief Introduction to DRT

1.1 Direct Value Theories vs. Logical Form Theories

1.1.1 The Direct Value Approach: Montague Grammar

• The opening lines of Montague's article 'Universal Grammar' are among the most often quoted ones within formal semantics:

"There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory." (Montague 1970)

Whatever we may think about this statement today, at the time there was an important reason for making a statement as uncompromisingly categorical as this. For arguably only such radical terms could have suitably conveyed the radically new and centrally important insight they expressed: that there are substantial fragments of natural language for which a syntax and model-theoretic semantics can be given in essentially the same form, and with the same mathematical precision, as what at the time (1970) had become the standard mode of presenting the syntax and semantics of formal languages like First Order and Higher Order Predicate Logic.

Montague's accomplishment was to show that and how this can be done, for fragments of English in which everything can be expressed that can be expressed in First Order Predicate Logic. And that, everyone with an interest in such matters was aware of this by that time, was pretty close to everything.

That significant fragments of natural languages could be described in this manner, and consistently with speakers' judgments about grammaticality and about the meanings of those expressions that are judged grammatical, was not simply new; it is was revolutionary. To the extent that people had thought about this possibility at all until then, they had been highly skeptical, either because they thought the project was too ambitious, or because they were convinced that the essence of what was human about human language would get lost underfoot; or some tangled combination of the two.

• Like the model-theoretic semantics specified by the standard presentations of Predicate Logic, the semantics in Montague's treatments of English fragments is a 'Direct Value Semantics':

Syntactically well-formed expressions are mapped directly on their semantic values (also: denotations) in models.

In the case of Predicate Logic this is done by a recursive truth definition (more precisely: a definition of truth on the basis of a recursive definition of satisfaction).

In Montague Grammar (MG), definitions of satisfaction (of formulas of Predicate Logic in models by variable assignments) directly relate the syntactic structures of the formulas to their semantic values in the models. These values are identified by means of 'Lambda terms', expressions from the *Typed Lambda Calculus*, a formal language that incorporates Predicate Logic but *has a much richer ontology.*

Montague needs this richer ontology because in his view there are natural expressions in his fragments that denote values which are not part of the ontology of First Order Predicate Logic. Two examples: (i) D(eterminer) P(hrases) like *every man* or *John*; (ii) determiners like *every.*

• Let us go into a little more into detail here. The ontology of the Typed Lambda Calculus is determined by its *types.*

In the classical *Extensional* Typed Lambda Calculus the types are generated from two base types, e (the type of individuals) and t (the type of truth values).

From these base types an infinite set of types can be generated by the principle that when α and β are types, then so is $<\alpha,\beta>$.

Types determine *Domains*. The Domains D_e and D_t of the types e and t, respectively are the set of individuals and the set $\{0,1\}$ of truth values (1 for 'True' and 0 for 'False'). We refer to the elements of D_e as *individuals*. Types of the form $<\alpha,\beta>$ are *function types*; the elements of the Domain $D_{<\alpha,\beta>}$ are functions that map the elements from the Domain D_α to elements of the Domain D_β.

Among the types that are part of the ontology of First Order Predicate Logic are: e, t, $<e,t>$, $<e,<e,t>>$, $<t,t>$, $<t,<t,t>>$. All these are needed in a formalism that can describe the semantic values that are needed in Montague Grammar. But other types are needed as well. Examples are $<<e,t>,t>$ (according to Montague the type of DPs like *every man* or *John*) and $<<e,t><<e,t>,t>>$ (according to Montague the type of Determiners, such as for instance the quantifier *every*.

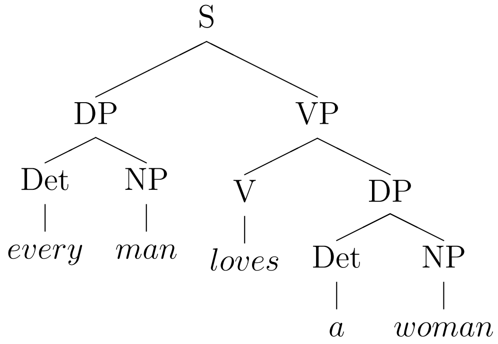
The models for the Extensional Typed Lambda Calculus cover this ontology: Each model M has Universes $D_{M,\alpha}$ for each of the types α, with the constraints that (i) $D_{M,t} = \{0,1\}$ and (ii) $D_{M,<\alpha,\beta>}$ is the set of functions from $D_{M,\alpha}$ to $D_{M,\beta}$.

• Here is a simplified illustration of Montague's model-theoretic semantics.

Consider the sentence

(1.1) Every man loves a woman

and suppose its syntactic structure is given by the following tree:



Each of the words of this sentence (and, generally, of well-formed expressions of the given English fragment that contains this sentence) is assigned a constant term of the Lambda Calculus, which must be of the right type.

Some of these, viz. *every* and *a*, are *logical terms*, they can be defined just in terms of the logical vocabulary of the Calculus. The others are 'atomic' constants. For some of these we may impose certain constraints on their values in models, thereby limiting the set of 'admissible' models for the given fragment. For instance, assuming that the fragment contains both the adjectives *dead* and *alive*, it is natural to adopt the constraint that no entity is both alive and dead at the same time. Such constraints on admissible models are known as *Meaning Postulates*.

The model-theoretic semantics now takes the following form:
(i) Each node in the tree (1.1) is assigned a lambda term.

(ii) The leaves are assigned the lambda terms associated with the words that are found there.

For each mother-daughters configuration in the tree a semantic operation is defined which combines the lambda terms for the daughters into a lambda term for the mother.

For instance for the configuration ' [[*every*].Det [*man*].NP].DP' there is a rule that combines the lambda terms τ for *every* (of type $<<e,t><<e,t>,t>>$) and σ for *man* (of type $<<e,t>,t>$) into the 'functional application term' $\tau(\sigma)$.

This term denotes in every model M for the Typed Lambda Calculus the result of applying the function denoted by τ to the entity denoted by σ.

• The Lambda Calculus that Montague himself makes use of is an *Intensional* formal system, which he called *Higher Order Intensional Logic* (HOIL). This system has an additional basic type s (for 'sense'), the role of which is more restricted than that of e and t. s makes it possible to turn each Domain D_α into a corresponding Domain $D_{<s,\alpha>}$ consisting of 'intensions of entities of type α': functions from a set W of 'possible worlds' to elements of D_α. Intuitive cases of this are (i) the Domain $D_{<s,t>}$ of *propositions*, (ii) the Domain $D_{<s,e>}$ of *individual concepts* and (iii) the Domain $D_{<s,<e,t>>}$ of *properties*. Some explanation:

(i) The Domain $D_{<s,\alpha>}$ consists of the functions from possible worlds to truth values. So the intensional value of a sentence S is the function which maps each possible world w to the truth value of S in w. These functions are plausible formal explications of the concept 'the proposition expressed by a sentence'. By virtue of what it means a sentence S will have usually varying truth values in different possible worlds about which it can be taken to make some claim, which can be either true or false. For those who think that what truth values S has in the different possible worlds provides an exhaustive picture of its meaning, the function from worlds w to S's truth values in w provides a natural formal analysis of 'meaning of S'; for those who think there is more to the meaning of S, the function can be at the very least a useful approximation.

(ii) The Domain $D_{<s,e>}$ is the Domain of functions from possible worlds to individuals. This Domain is identified with the domain of *individual concepts*. For an example consider for instance the definite description *the first man to run the mile in less than 4 minutes*. In our actual world the individual denoted by this description was the English athlete Roger Bannister. The individual concept denoted by the description is the function which maps each possible world w to the one who was the first in w to run the mile in under 4 minutes (and which maps worlds in which the mile was never run in under 4 minutes by any one some default value #).

(iii) The Domain $D_{<s,<e,t>>}$ of *properties*: a 1-place predicate P of individuals has an extension, the set consisting of all the individuals of which the predicate is true; but its extensions usually vary from world to world, and what their extensions are in the different worlds w is a reflection of the 'meaning' of P. Here too we can distinguish, as we did under (i), between those for whom the extensions of P in the different worlds exhaustively capture the meaning of P and those for whom there is more to its meaning. For the former the function which maps each w to the extension of P in w will be an exhaustive analysis of the property expressed by P; for the latter it will be a useful approximation.

Adding s as a basic type has repercussions throughout the type system. The simplest way to modify the system is to add the rule that whenever α is a type, then so is $<$s,$\alpha>$, and that $D_{<s,\alpha>}$ is the set of functions that map the set of possible worlds to entities from $D_{<s,\alpha>}$.[1]

[1]There are alternatives to HOIL in which intensionality is handled differently. On the one hand, in the so-called *Type 2* system of (Gallin 1975), in which possible worlds are

The models for this intensional version of the Typed Lambda Calculus must have a richer ontology than those for the Extensional Typed Lambda Calculus. A model M must have in addition to the Domains of models for the Extensional Typed Lambda Calculus (a) a set W of 'possible worlds', and (b) for each of its Domains D_α it must have a corresponding Domain $D_{<s,\alpha>}$.

• In this outline of HOIL we have been speaking throughout of a set of 'possible worlds', without saying much about its members. In applications it is often assumed that these members are not just possible worlds in an intuitive sense – that of 'other ways the world could have been' – but something else (for instance times), or something more complex (combinations of worlds and times). To make clear that the members of the set need not always be possible worlds, it is common to talk about 'indices' rather than 'worlds' in presentations of HOIL and other intensional versions of the Typed Lambda Calculus. Applications in which the set consists of pairs $<w,t>$ of worlds and times are of particular importance for linguistics. We will also assume indices of this form throughout this introduction.

• A crucial feature of direct value semantics in the sense of Montague is that the operations defined for the mother-daughters configurations are always **operations on the denotations of the lambda terms**, and not on the terms themselves, as expressions with their own syntactic structure. The operations never make an essential use of syntactic properties of the terms.

The lambda terms that are associated with syntactic nodes by a direct value semantics theory like MG can be referred to as 'Logical Forms' and sometimes are. There can be no objection to this so long as we see it as a mere matter of terminology. Crucial to the direct value approach is that the compositional operations that are used in the compositional determination of the semantic values for complex expressions are always operations on the values

treated as a special kind of individuals, which stand to other entities from the Lambda Calculus in certain special relations such as for instance that of the proposition expressed by a sentence S being true in a world w. Type 2 is thus a kind of *two-sorted* Lambda Calculus, in which a certain Domain is sub-divided into different 'sorts', here the Domain D_e into the sort of worlds and the sort of other individuals. Type 2 is often used by Linguists because it simplifies certain aspects of notation and of the conceptualization behind it.
In the opposite direction it is also possible for purposes of linguistic analysis to make do with a more restricted set of types, by allowing only types of the form $<$s,$\alpha>$, when α is a type from the extensional Typed Lambda Calculus (thereby avoiding types like $<$s,$<$s,$\alpha>>$, for which it is hard to find independent linguistic motivation.

denoted by the lambda terms, not operations on the terms themselves, as syntactic objects.

The direct value approach has been the dominant one in formal semantics since Montague's path-breaking contributions in the late sixties and early seventies. But there are linguistic phenomena that MG is not properly geared to handle. Some of these difficulties came into focus in the late seventies. Discourse Representation Theory was one proposal to deal with them. DRT is an instance of the 'Logical Form' approach to natural language semantics, in which Logical Forms of natural language sentences and discourses are computed from syntactic parses. The model-theoretic semantics is then given for these Logical Forms, and thereby also, albeit indirectly, for the natural language sentences and discourses from which the Logical Forms are derived.

Since DRT provides the basis for MSDRT, which shares many properties with it, some things should be said about DRT first. In an introduction to MSDRT they would have to be addressed at some point or other anyway, and to present them as the general properties of DRT they are is a way of making clear that these are features that MDRT also shares with other extensions of DRT.

1.1.2 A First Challenge to MG: Tense in Discourse; the origins of DRT

• The original motivation for DRT had to do with the way in which natural languages (French, English, German ...) handle reference to time.

A primary role in the ways many natural languages deal with temporality is played by the tenses of their verbs. In many languages tense forms do not only carry information about the temporal locations of the situations or episodes that are being described – i.e. where these are in relation to the current time (or utterance time) or some other orientation time – at that time, in the past of it or in the future of it.

But to think of tenses as *only* serving location in time. Some also carry information about *aspect*.

A well-known example are the French tense forms *Imparfait* and *Passé Simple*.

A similar contrast is found in English between simple past and past progressive.

The aspectual properties that are encoded in these tenses manifest themselves in part through interaction with another salient feature that tenses have more generally: that of relating the states or events described by their clauses temporally to adjacent sentences. Thus the aspectual differences between *Passé Simple* and *Imparfait* manifest themselves prominently when they occur in sentences that are part of an ongoing discourse. An example is the contrasting pair (1.2.a) and (1.2.b).

(1.2) a. Lorsque Jean ouvrit les yeux, il vit sa femme, qui était assise près de son lit, Elle luit sourit.
When Jean opened his eyes he saw his wife, who was sitting by his bedside. She smiled at him.

b. Lorsque Jean ouvrit les yeux, il vit sa femme, qui était assise près de son lit, Elle luit souriait.
When Jean opened his eyes he saw his wife, who was sitting by his bedside. She was smiling at him.

Another point about the aspectual dimension of Passé Simple and Imparfait is an intuition that has been expressed by various scholars of the Grammar of French. It is that:

Sentences with verbs in the Passé Simple tend to express 'punctual' events. Sentences with verbs in the Imparfait tend to express temporally extended events or states.

To the extent that this second observation is correct, it is best explained in terms of the *discourse time*. The discourse time of a given multi-sentence piece of discourse or text – most typically narrative texts, but we find this also in other text types – is a temporal structure that is immanent to the discourse or text itself. It is a structure that can be constructed from the events and states introduced by the sentences and clauses of the text, among them, and in particular, its sentences and clauses in the Passé Simple and the Imparfait.

Because these two tenses connect their sentences differently to the sentences preceding them, the events that are contributed to the discourse representation by Passé simple sentences typically reemerge as punctual elements (temporal instants) of the discourse time while the events or states that are introduced by Imparfait clauses and sentences tend not to be punctual in this sense. But this is punctuality and non-punctuality in a different sense from that of the time of the actual of fictional world that the text or discourse is about. There is no contradiction between an event that is mentioned in the discourse or text being punctual in the sense of the discourse time while having temporal extension in the world described.

Here is a little bit of the formal background to the punctuality and durativity of events in states in discourse time and in real time.

Suppose that an ordered structure of eventualities (events and states) satisfied the following postulates. ($\prec$ stands for complete temporal precedence; O for temporal overlap.)

(A1) $(\forall x)(\forall y)(\, x \prec y \rightarrow \neg\, y \prec x)$
(A2) $(\forall x)(\forall y)(\forall z)(\, x \prec y \;\&\; y \prec z \rightarrow x \prec z)$
(A3) $(\forall x)\; x \,\mathrm{O}\, x$
(A4) $(\forall x)(\forall y)\; x \,\mathrm{O}\, y \rightarrow y \,\mathrm{O}\, x)$
(A5) $(\forall x)(\forall y)(\, x \prec y \rightarrow \neg\, x \,\mathrm{O}\, y)$
(A6) $(\forall x)(\forall y)(\forall z)(\forall t)(\, x \prec y \;\&\; y \,\mathrm{O}\, z \;\;\&\;\; z \prec t \rightarrow x \prec t)$
(A7) $(\forall x)(\forall y)(\, x \prec y \vee\; x \,\mathrm{O}\, y \vee y \prec x)$

• Bertrand Russell and Norbert Wiener (who was Russell's assistant at the time when they jointly did this work) showed the following: When a structure $<EV,\prec,\mathrm{O}>$ of eventualities[2] satisfies these axioms, then it is possible to define an ordering of temporal 'instants' on the basis of it. The instants of this ordering are maximal sets of overlapping events and states. That is:
T (the set of instants) = $\{\mathrm{t}\colon \mathrm{t} \subseteq \mathcal{P}(EV) \;\&\; (\forall ev)(\forall ev')(ev \in \mathrm{t} \;\&\; ev' \in \mathrm{t} \rightarrow ev \,\mathrm{O}\, ev') \;\&\; \neg(\exists ev'')(ev'' \in EV \;\&\; (\forall ev)(ev \in \mathrm{t} \;\&\; ev'' \,\mathrm{O}\, ev)) \;\&\; \neg\, (ev'' \in \mathrm{t})\}$

Let furthermore the relation $<$ be defined by: $\mathrm{t} < \mathrm{t}'$ iff $(\exists ev \in \mathrm{t})(\exists ev' \in \mathrm{t}')\; ev \prec ev'$.

It is not hard to show that when $<EV,\prec,\mathrm{O}>$ satisfies Axioms (A1) - (A7), then $<\mathrm{T},<>$ is a strict linear ordering (that is, $<$ is transitive, asymmetric and total).

The instants t of T are naturally thought of as the sets of those and only

[2] From now on we follow the practice of using the term *eventuality* as a cover term for both events snd states.

those eventualities from EV that go on at them – ev goes on at t iff $ev \in$ t.

So far all this is Russell-Wiener. We now come to the application of their construction to events and discourse time. This takes a little bit of unpacking. But the unpacking will be useful also for what will have to be said about DRT in the next section.

The examples in (1.2) give an impression of what contributions to the overall meanings of these two-sentence texts are made by the Passé Simple and Imparfait. The challenge is to turn those intuitions into precisely formulated rules that capture how these tenses are part of introducing new events and states into the discourse and how these are temporally related to other events and states that have been introduced already. And that challenge is part of a bigger one: to formulate such rules not only for the tenses but also for all other grammatical constituents of the text, so that application of them also captures the contributions those other constituents make to the meaning of the text.

It was an attempt to meet those challenges that gave rise to DRT. Its first version took the form of formulating rules for all sentence constituents of some certain narrative texts – quite simple ones to start with, but with interpretation rules for a few tense forms including Passé Simple and Imparfait as well as simple predicational constructions with verbal and nominal predicates. But the formulation of such rules entails certain structures that come about when they are applied. These structures must truthfully represent the content of the texts from which they are constructed. In other words, they must be ascertainable as correct representations of the texts they represent. To that end DRT treats these representations as 'formulas' of certain formal languages – its *Logical Form Languages* or *DRS Languages*, where *DRS* is short for 'Discourse Representation Structure', the name of the semantic representations that DRT constructs form narrative texts and other types of texts and discourses. The DRS languages are specified independently of how DRSs are constructed, by defining their syntax and, building on that, a model-theoretic semantics for them (much like we nowadays do in presentations of the Predicate Calculus). The check on whether the construction rules do the right thing when they are applied to the texts and discourses within their scope is then whether the truth conditions that the model theory for the given DRS language assigns to the resulting DRS match our intuitions about the truth conditions of the represented discourse or text (Kamp 1981*a*). (For the record: The hard part in this was formulating the interpretation rules for the tenses and other sentence constituents. The form of the representations, the syntax of the DRS languages to which those belong and the model theory for those languages, came along and fell into place as a kind of secondary effect.)

Detailed descriptions of the different components of DRT – its construction algorithm (the set of interpretation rules and their application and the syntax and semantics of DRS languages – can be found in the literature.[3] Some more about the construction of DRSs can be found in the next section. But here our focus is only on the temporal aspects of the DRSs that a construction algorithm builds from the sentences and texts for which it is designed. The rules of those algorithms for tenses and certain other constituents do two things that are relevant to the temporal structure of the constructed DRS: (i) they introduce events sand states into the DRS and (ii) they establish temporal relations – $\prec$ for temporal precedence, O for temporal overlap, $\subseteq$ for temporal inclusion, $\supset\subset$ for temporal abutment – between these events and states. This temporal structure, consisting of the events and states from the DRS and the temporal relations between them, can be extracted from the DRS by eliminating all else. We refer to this structure as the *event structure from* the given DRS. As a rule this event structure will be in essence like the event structures of Russell and Wiener and the Russell-Wiener construction will convert it into a linearly ordered instant structure. (In this application of the Russell-Wiener construction we make no distinction between events and states; the states of the event structure count as events for the construction in the same way that the events of the event structure do. Also the additional temporal relations $\subseteq$ and $\supset\subset$ make no significant change to the construction, something that is easily verified and left to the reader.) Finally, as we will see in the next section, DRS Construction Algorithms produce DRSs with times as well as events snd states. These too should be included in the event structures that can be extracted from those DRSs. In the application of the Russell-Wiener construction to such event structures the times also should be treated as events.) This linear instant structure is called the *discourse time* of the discourse from which it is computed via DRS and event structure.

For narrative and other texts with a prominent presence of Passé Simple and Imparfait sentences, and especially when these are the only tenses occurring in them, the instant structure derived from it will have the property that the events that were introduced by the Passé Simple occupy single instants from

[3] Most details can be found in (Kamp & Reyle 1993). The 'top down' construction algorithm used throughout this book is no longer used, we believe, by anyone. The most extensive 'bottom up' algorithm, in which DRSs are built by a transparently compositional process from the leaves of the syntactic sentence trees is as yet unpublished (Kamp 2021*a*).

the discourse time. This is the sense in which such events are punctual.

The model theory for DRT is from an abstract formal point of view much like the standard model theory for the Classical Predicate Calculus with a definition of the truth values of DRSs in models that is recursive with respect to the syntactic structure of the DRS. But there is a way of looking at the truth relation between DRSs and models that is captured by this definition that doesn't apply to all DRSs but does to DRSs for paradigmatic examples of simple narrative texts. On this view the truth relation is one of *faithful embeddability* of the DRS into the model: the DRS is true in the model iff there is a faithful embedding of it in the model. A faithful embedding f for a DRS K in a model M – the more common term used in DRT is *verifying embedding* – maps the entities from the DRS to entities of the model in a manner that preserves the constraints that the DRS imposes on its entities (by means of its *DRS Conditions*, see the next section). This includes the temporal relations $\prec$, O, $\subseteq$ and $\supset\subset$. So a verifying embedding will also be a verification of the event structure that can be extracted from the DRS. But note that such verifications do not have to map events from the DRS that are punctual in the discourse time onto events from the model that are punctual in the sense of the time of the model; the latter may be temporally extended. The reason can be most easily seen for the common case of a verifying embedding f that is an *injection* of the event structure into the model (mapping different events and states of the event structure to different events and states in the model). In such cases the restriction of f to the map f_{ev} that is defined only on the events and states of the event structure can be reversed into a map f_{ev}^{-1} from the model onto the event structure. And this map can be seen as a kind of contraction of the time of the model to the discourse time (formally: a homomorphism with respect to the temporal relations), where extended periods from the former (among them the durations of events $f(e)$ in terms of the time of the model) can be mapped to single points of the discourse time.

In this sense an event that is described by a Passé Simple sentence and introduced by that sentence into the discourse can have temporal extension (as pretty much every event will that we can think and talk about) and nevertheless it can be punctual in the discourse-related sense that is perceived by those who put forward the claim that Passé Simple events are punctual.

The puzzle that has occupied us in these last few pages is how it is possible for speakers to have a strong intuition that events described by Passé Simple sentences are punctual even when they would acknowledge that the events spoken have non-zero extension in physical time. Our explanation rests on the assumption that the 'punctuality' of such events must be in the sense of some other concept of time; and obviously that concept must have to do with the way these events are described, in the context of the discourse of which their description occurs, rather than with the description-independent properties of the event. (And it is hard to see, we think, what other form an explanation of this could take.) It seems reasonable to refer to this concept of time as the 'discourse time'. But what precisely is it, and how is it determined by the discourses in which the Passé Simple makes the contributions that lead to the intuition that the events it introduces are punctual in the sense of the discourse time? These are the questions that DRT was set up to answer and that could only be achieved, it seemed, by describing in detail how the interpretation of tenses and other sentence constituents lead to semantic representations from which the discourse time can be derived by a general procedure. So DRT was driven inexorably towards its Logical Form architecture, in which Logical forms not only determine truth conditions but also provide the basis for a discourse time in which Passé Simple events are demonstrably punctual.[4]

• The account of the temporal and aspectual dimensions of certain tense forms that has just been rehearsed points to the following conclusions:

1. There are systematic aspects of natural language that have to be handled at the level of multi-sentence discourse, and not at that of the single sentence.

2. Some phenomena seem hard to account for without including eventualities in our ontology.

3. The account we have sketched of what supports the intuition that Passé Simple events are punctual depends crucially on the concept of a semantic representation of a discourse, which defines its own time structure and is distinct both form the actual discourse from which it is derived through interpretation and from the models in which natural language sentences and discourses are assumed to be true or false. This suggests that human interpreters build discourse representations with the kind of event structure that is assumed by the reconstruction.

4. In the construction of these semantic representations the representation

[4]For more discussion of the event structures that are established through discourse representation see (Partee 1984) and (Kamp 2019).

of that part of the discourse that has been interpreted so far plays a double role: (i) capturing the content of what has been so far interpreted and (ii) serving as context for the interpretation of what comes next. In DRT this double function of its semantic representations (its DRSs) is crucial. One of the criteria of success in applications of DRT is that the DRSs which are constructed incrementally as the discourse is traversed from left to right can do both these things at once.

It is unclear how Direct Value accounts of truth conditions, which restrict attention to syntactic sentence structure and models would be able to account for such phenomena.

The discourse behavior of tenses is one reason for wanting Semantic Representations – or Logical Forms; we will be using these terms as interchangeable – as well as a model-theoretic semantics for natural language. And when there are independent reasons for wanting Logical Forms, then it proves to be natural to do the model-theoretic semantics via the Logical Forms, and thus to adopt the Logical Form approach. But there are other reasons for going this way of going about natural language semantics as well. These too have a connection with multi-sentence discourse. But their main point is a difficulty with the way in which a direct value approach like MG identifies propositional contents expressed in natural language. Some of the examples that show this are multi-sentence discourses. But the problem also arises with individual sentences in isolation, and it was there that it was recognized first. This is the problem mostly referred to as 'donkey anaphora'.

1.2 More on DRT

1.2.1 Pronominal Anaphora

Consider the following pair of two-sentence texts. (Partee 1971)

(1.3) a. Exactly one of the ten marbles is not in the bag. It is under the sofa.

b. Exactly nine of the ten marbles are in the bag. It is under the sofa.

• The point of this example: According to standard assumptions of Natural Language Semantics the first sentences of (1.3.a) and (1.3.b) express the same proposition. But there is a strong intuition that in (1.3.a) the pronoun *it* can be understood as referring to the missing marble, but that that is not possible in (1.3.b).[5]

Ergo: Pronominal anaphora is sensitive to aspects of the forms of sentences and texts that make no difference to their truth conditions (viz. to the propositions expressed, thought of as sets of possible worlds).

• Note also that this is a phenomenon that has to do with pronouns, as distinct from other anaphoric noun phrases. We do not find the same difference when we replace *it* in (1.3.a,b) by e.g. *the missing marble. the missing marble* does just as well in (1.5.b) as it does in (1.5.a).

(1.5) a. Exactly one of the ten marbles is not in the bag. The missing marble is under the sofa.

b. Exactly nine of the ten marbles are in the bag. The missing marble is under the sofa.

• Conclusion: A theory of discourse interpretation needs to keep track of more structure than just standard semantic values (such as propositions in the sense of sets of possible worlds).

[5] There have been extensive debates over how strong the contrast is between the acceptability of *it* in (1.3.a) and (1.3.b). For one thing, (1.3.b) becomes much better when there is a switch of turns between first and second sentence: the first is uttered by A addressing B and the second by B as a reply to the first sentence. And a longer pause between first and second sentence, even when uttered by the same person, makes the use of *it* more acceptable too. Perhaps this is because in these settings the use of *it* can be understood as *deictic* rather than anaphoric: the speaker uses *it* to refer to an entity that is salient in her mind, rather than as coreferring with a noun phrase occurring in an accessible position in the spoken or written discourse. On the whole the contrast between (1.3.a) and (1.3.b) seems quite robust, however, and many examples of similarly contrasting pairs have been contributed to the discussion. One in which the contrast is very strong for us is (1.4).

(1.4) a. One of the twenty people who set of for the hike didn't make it to the top. She sprained her ankle.

b. Nineteen of the twenty people who set of for the hike made it to the top. She sprained her ankle.

An interesting twist to this pair is that in (1.4.a) an interpreter who knows nothing about the event reported will without difficulty accommodate that the person who didn't make it to the top was a woman. But that accommodation is of no help in (1.4.b).

1.2.2 Temporal Anaphora and Aspect; DRS Construction

The two examples in (1.2) gave a first sense how the aspectual properties of tenses can manifest themselves through the ways in which their clauses are connected with the clauses and sentences surrounding them. These interclausal connections can be seen as the manifestations of principles of *temporal anaphora*: the tenses of finite clauses want to be connected with times or events elsewhere in the sentence or discourse context. These connections often reveal aspectual differences between the tense forms that establish them. Below we will see how DRT provides the structure that is needed to account for such difference. To that end we have a close look at another set of examples, shown in (1.6) (from (Kamp, van Genabith & Reyle 2011)), for which we will construct the DRSs that DRT proposes as their Logical Forms.

(1.6) a. Joseph turned around. The man was pulling a gun from his belt.

b. Joseph turned around. The man pulled a gun from from his belt.

c. Joseph turned around. The man had pulled a gun from his belt. He was pointing it at Joseph.

• (1.6.a) and (1.6.b) are another instance of the contrast we saw in Section 1.2. (Alain who opens his eyes and sees his wife.)

The first two sentences of (1.6.c) are a further variation of this theme: The past perfect of the second sentence locates the event described by its sentence in the past of the event described by the first sentence.

The third sentence of (1.6.c) contains instances of both nominal and temporal anaphora: the pronoun *it* refers back to the gun mentioned in the second sentence and the state of pointing the gun at Joseph is presented as going on when Joseph turned around.

• We focus on (1.6.c). We first show what the DRSs for the sentences of (1.6.c) look like and then how these DRSs can be constructed from syntactic parses for the sentences.

(1.7) (DRS for 1st sentence of (1.6.c))

$$\begin{array}{|c|}\hline e \quad j \\ \hline e \prec n \quad \text{Joseph'}(j) \\ e\text{: turn-around'}(j) \\ \hline \end{array}$$

Comments:

1. We use primes in the DRS predicates that translate natural language predicates, as in *turn-around'*.

2. A DRS consists of two parts, its *Universe* and its *Condition Set*.
The elements of DRS Universes are *discourse referents* (or *drefs* for short).
The Conditions in Condition Sets are simple or complex predications.

3. Verbs and their projections are treated as descriptions of *eventualities*, which for us will always be either *events* or *states*.

So, an intransitive verb like *turn-around* has not one but two arguments, its *referential argument* – the described eventuality, here an event e – and the non-referential 'subject argument' Joseph.

4. The eventualities described by verbs are located in time by tense and sometimes additionally by temporal adverbials (like *yesterday*, *next week*, *on the first of February*, *after the party* and so on). In (1.7) the simple past tense locates e in the past of the utterance time n.

5. The Condition 'Joseph'(j)' expresses that j represents the referent of the name *Joseph* as used by the producer of (1.6.c). (This is a stop-gap measure. For a more lifelike treatment of proper names see e.g. (Kamp 2015).)

(1.8) (DRS for 2nd sentence of (1.6.c))

e' m g b u
$e' \prec n$ $e' \prec e$ [*the man*](m) $u = m$ gun'(g) belt'(b) POSS(u,b) e': pull-from'(u,g,b)

Comments:

1. One way in which (1.8) differs from (1.7) is in having an *undeclared dref*, viz. e, which occurs in the Condition '$e' \prec e$'. It has been imported from (1.7) as part of the anaphora resolution of the past perfect *had pulled*. As we will see below, undeclared discourse referents in DRSs prevent those DRSs from expressing propositions on their own. They can only contribute to the expression of a proposition when they are 'merged' with the DRSs from which

their undeclared drefs are imported.

2. The construction of (1.8) also involves a case of sentence-internal anaphora: the pronoun *his* is anaphoric to *the man*. This anaphoric connection is expressed in (1.8) by the equation '$u = m$', where m is the dref introduced by the noun phrase *the man* and u the dref introduced by the pronoun *his* (see the DRS construction for this sentence that is shown below).

3. (1.8) illustrates an important difference between pronoun anaphora and the temporal anaphora triggered by tenses. The semantics of pronoun anaphora is always coreference: an anaphoric pronoun always takes over the referent from its antecedent. For temporal reference this is not so. For instance the past perfect of the second sentence of (1.6.c) locates the event described by its verb in the past of the temporal antecedent, viz. the event described in the first sentence, a relationship represented in (1.8) by the Condition '$e \prec e'$'. (Other tenses express other relations between the eventualities they locate and the 'antecedents' that they relate those eventualities to.)

4. (1.8) too involves a stopgap treatment of a definite noun phrase, this time the definite description viz. *the man*. The Condition '[*the man*](m)' is used to express that m represents the referent of the noun phrase *the man*. Definite descriptions are the most versatile of the different types of definite noun phrases of English, and a proper treatment of them correspondingly complex.[6]

5. The predicate 'POSS' is a portmanteau predicate for all the different relations that can be expressed by 'possessive vocabulary', including genitival noun phrases, possessive pronouns, the preposition *of*, the verb *have*. The question what kinds of possessive relations there are and the question how we recognize which tokens of possessive constructions express which of these relations are notoriously hard.[7]

[6] One important distinction between anaphoric definite descriptions – anaphoric in the sense that their interpretation involves some earlier part of the sentence or discourse – is that between those descriptions that select their referents from some set made available by the antecedent discourse or some other part of the sentence (Strategy 2) and those which pick up as their referent an entity previously introduced into the discourse (Strategy 1). Strategy 1 descriptions compete with anaphoric pronouns (and are often preferred when a pronoun would be ambiguous between the intended antecedent and some other one. The distinction between Strategy 1 and Strategy 2 descriptions has connection with that between referential and attributive descriptions in the sense of Donnellan, but the two distinctions are not the same (for discussion see (Kamp 2022*b*)).

[7] For a detailed study of possessive descriptions see (Barker 1995) (**Other references**

• After the DRS (1.8) has been constructed (in part with the help of its 'discourse context' (1.7)), it can be merged with its discourse context, leading to the semantic representation for the two sentences together:

(1.9) (DRS for the 1st and 2nd sentence of (1.6.c))

$$
\begin{array}{|c|}
\hline
e \quad j \quad e' \quad m \quad g \quad b \quad u \\
\hline
e \prec n \quad e' \prec n \quad e' \prec e \\
\text{Joseph'}(j) \quad [\textit{the man}](m) \quad u = m \\
\text{gun'}(g) \quad \text{belt'}(b) \quad \text{POSS}(u,b) \\
\\
e\text{: turn-around'}(j) \\
e'\text{: pull-from'}(m,g,b) \\
\hline
\end{array}
$$

(1.10) is the DRS for the 3rd sentence of (1.6.c).

$$
(1.10)\;
\begin{array}{|c|}
\hline
s \quad u' \quad g' \quad j' \\
\hline
s \prec n \quad e \subseteq s \\
\text{Joseph'}(j') \quad u' = m \quad g' = g \\
\\
s\text{: PROG}(\hat{}e.
\begin{array}{|c|}
\hline
e \\
\hline
\\
e\text{: point-at'}(u`,g',j') \\
\hline
\end{array}
\;) \\
\hline
\end{array}
$$

<u>Comments</u>:

1. This sentence is multiply connected with its discourse context (the DRS (1.9)):

(i) via the equation '$z = m$' (resolution of *he*);

(ii) the equation '$v = g$' (resolution of *it*); and

(iii) the temporal inclusion Condition '$e \subseteq s$' (resolution of the 'progressive past tense *was pointing*.)

---

here?).

2. The treatment of progressive tenses in DRT assumes that the progressive is an operator that turns an event description into the description of a state that consists in the going on of a process that will or could lead to a complete event of the kind described by the non-progressive form of the verb or verb phrase. (A completed event need not be the actual result of the process. The details of the way in which progressives are treated and represented in DRT are somewhat complex, but do not matter in the present context. For important proposals about the semantics of the English progressive see (Dowty 1979), (Landman 1989*b*).)

• Once again merging the DRS (1.10) for the third sentence with its discourse context (1.9) gives the DRS for the full three sentence discourse.

(1.11)(DRS for (1.6.c))

$$
\begin{array}{|c|}
\hline
e \quad j \quad e' \quad m \quad g \quad b \quad s \quad z \quad v \quad j' \\
\hline
e \prec n \quad e' \prec n \quad e' \prec e \quad s \prec n \quad e \subseteq s \\
\\
\text{Joseph'}(j) \quad [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \\
\text{POSS}(u,b) \\
\\
e\text{: turn-around'}(j) \\
e'\text{: pull-from'}(m,g,b) \\
\\
s\text{: PROG}({}^\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: point-at'}(m,g,j') \\ \hline \end{array}) \\
\hline
\end{array}
$$ [8]

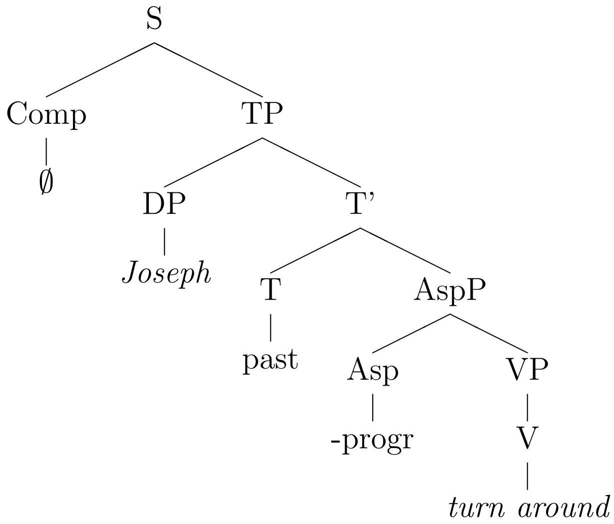
• We now show how these DRSs can be constructed from syntactic parses for the three sentences of (1.6.c).

(1.12) gives the syntactic parse we assume for the first sentence.

---

[8]The truth conditions represented by the Condition ‘s: PROG(${}^\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: point-at'}(m,g,j') \\ \hline \end{array}$)’ are fairly complex; they are intensional in that they involve the*property* of being an event e of z pointing v at j'. I do not go into details aboiut this, since it is of no direct relevance to the points that matter here. For details see (Kamp et al. 2011).

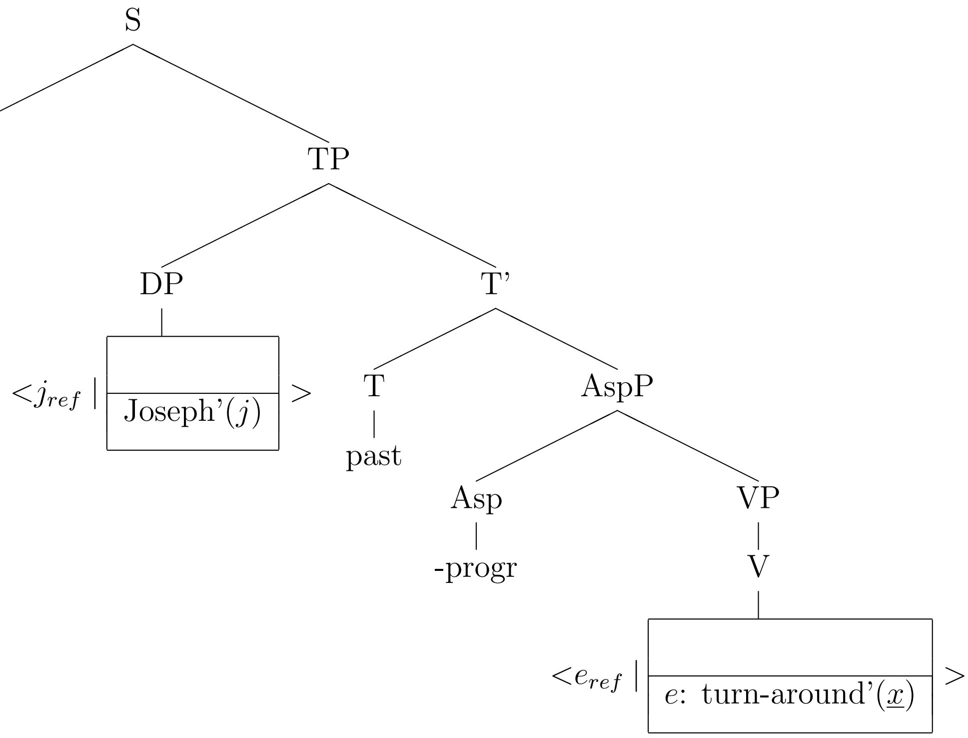
(1.12)



The DRS for the first sentence can be constructed from (1.11) 'bottom up' as follows:

We start with *lexical insertion* for the words occurring in (1.11). This gives us (1.13).

(1.13)



Comments:

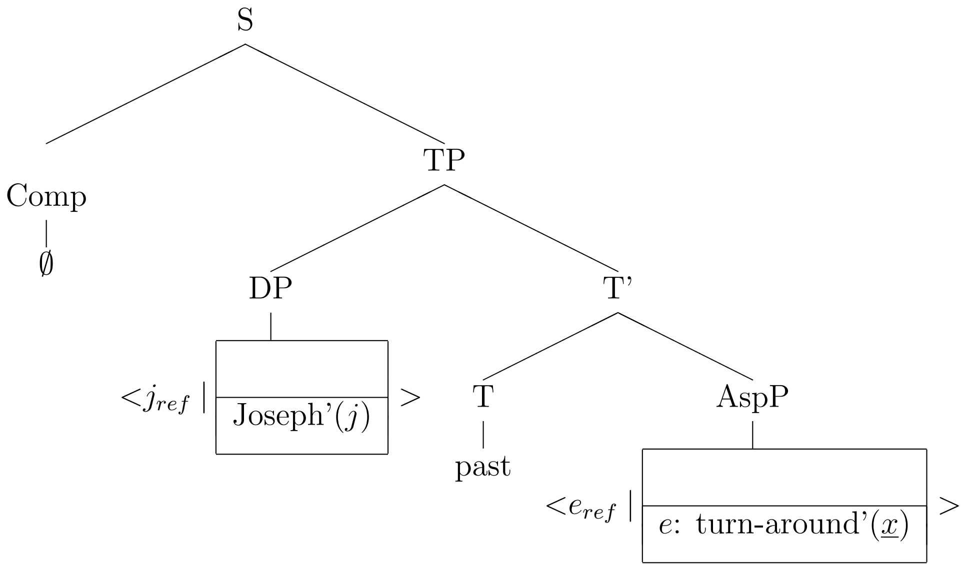
1. The version of DRT we are using here distinguishes between argument terms, which are always drefs, and *argument slot symbols*, like the symbol '$\underline{x}$' that occupies the *non-referential argument position* of 'turn-around' in (1.13). The distinction between *referential arguments* of predicates and their *non-referential arguments* is that non-referential arguments are normally realized by separate phrases, whereas their referential arguments never are. (Example: the 'relational' noun *friend* has two argument slots, one for its referential argument and one for its non-referential argument. In the phrase *friend of Maria* the non-referential argument slot is filled by the noun phrase *Maria*. There is no separate phrase for the referential argument.) The referential argument of a natural language predicate word is an inseparable part of the predicate itself. In our version of DRT this is formalized as the principle that the lexical semantic representation of a predicate word comes with a dref that plays the part of its referential argument.This dref is part of what gets inserted when occurrences of the predicate word in syntactic trees are replaced by their semantic representations.

For verbs the referential argument is always an *eventuality* (either an *event* or a *state*). Example: the intransitive verb *turn around*. Intransitive verbs have besides their referential argument – for the eventuality the verb is used to describe – one non-referential argument, which in sentences in the active mode is realized by the grammatical subject. (In the first sentence of (1.6.c) this is the subject phrase *Joseph*.) In the lexical semantic representation for *turn around* that is assumed in (1.13) its non-referential argument is marked by the slot $\underline{x}$. This slot will be filled by the referential argument of the subjective phrase. The referential argument of the subject *Joseph* will be inserted in the slot $\underline{x}$ when the semantic representation of the T' node of (1.13) is combined with that of the subject phrase.

2. The semantic representations for intermediate nodes of our syntactic structures (those below the D node), are not DRSs, but DRSs preceded by a *store*. The store is for drefs that have been introduced but must remain available for later *binding*. In DRT binding takes the form of placing the dref in some DRS Universe, while removing it from the store, if that is where it was. When all drefs in a store have been removed from it, the store is eliminated and the representation consists just of a DRS, as in (1.7) – (1.11). Occurrences of referential arguments in stores are usually marked with the subscript $_{ref}$. The principles for $_{ref}$-marking are somewhat complex and not discussed here.

We continue the representation construction for the first sentence of (1.6.c). The V representation is passed up unchanged to the VP node. The feature '-prog' of the Asp node signifies that the VP representation is passed up unchanged once more to the AspP node:

(1.14)



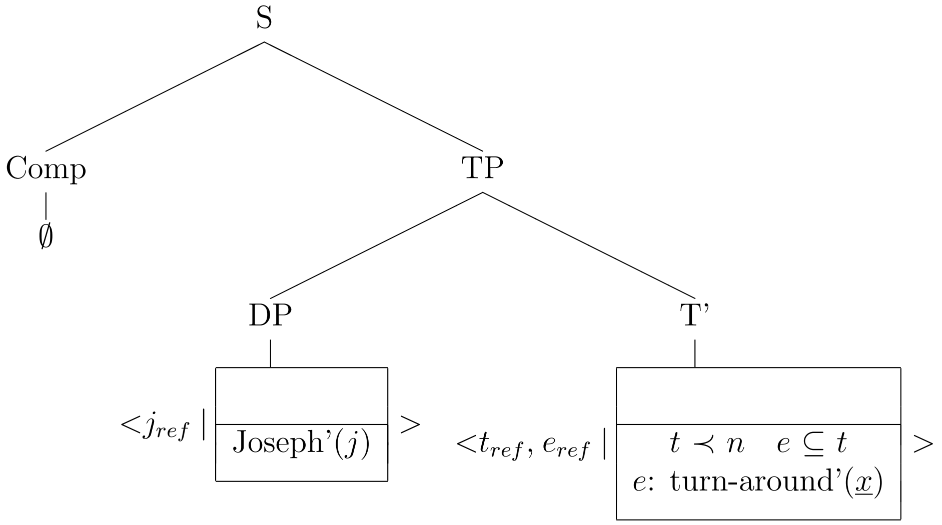
At this point the Tense feature 'past' makes its contribution to the semantic representation. In general, Tense features do three things:

(i) They locate the referential eventuality argument of the verb they govern by introducing a time dref, like t in (1.15) – the *location time* of the eventuality argument – and temporally relate this time 'anaphorically' to some other time in the discourse context. Mostly, as in this case, this other time is the utterance time n, but not always.

(ii) They tie the eventuality argument of the governed verb 'anaphorically' to some other eventuality or time in the discourse context. This is the 'temporal anaphora' of tenses. (For discourse-initial sentences, for which the discourse context is empty, this second aspect of tense interpretation evidently plays no part.)

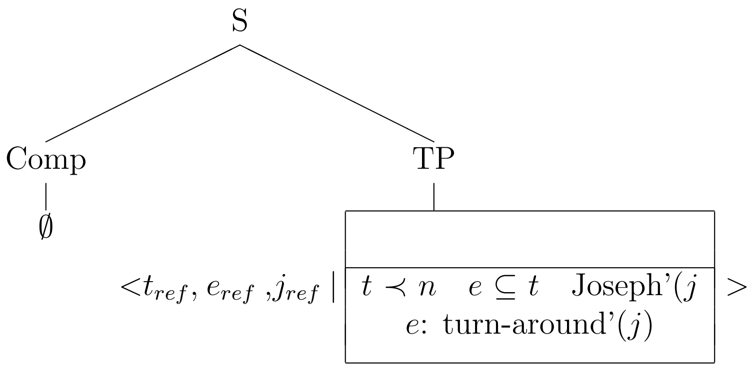
(iii) They connect eventuality argument to its temporal location time in one of two ways, depending on the *aspect* imposed by the way the clause describes the eventuality argument of the verb. DRT makes use of a binary aspectual distinction here, between events and states, assuming that it is generally possible to recognize from the properties of eventuality descriptions whether they describe events or states. Whether the eventuality ev is an event or a state determines its connection with its location time t. If ev is an event e, then the relation is '$e \subseteq t$'; if ev is a state s, then the relation is '$t \subseteq s$'.

(1.15)



The next step is a case of argument insertion: insert the referential argument j of the argument phrase – here the (D(eterminer) P(hrase)) *Joseph* – for its slot $\underline{x}$ in the sister representation. After the actual argument insertion the stores and the DRSs of the two sister representations are merged.

(1.16)

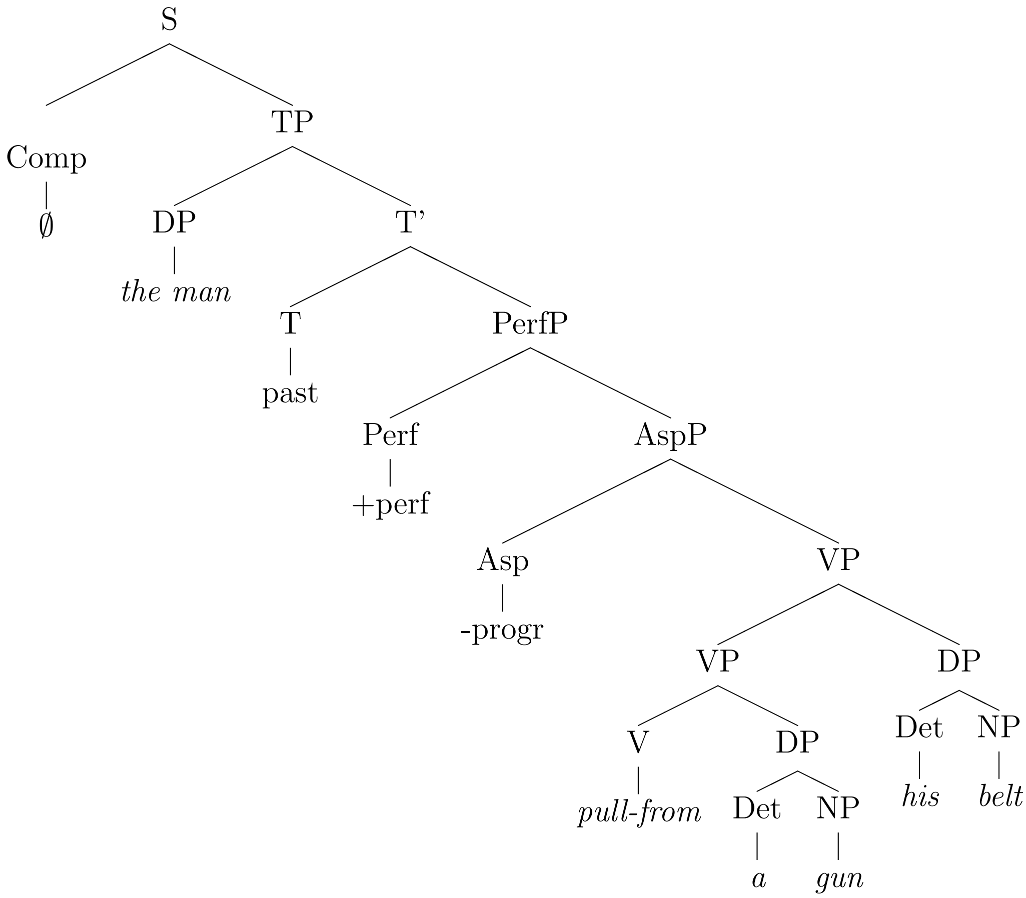


The final step is *existential binding* of the remaining drefs in the store, by transferring them to the Universe of the DRS following it, and elimination of the now empty store from the display of the representation. This gives the DRS shown in (1.17). We can eliminate t from this DRS without change in truth conditions, replacing the pair of Conditions ' $t \prec n$' and '$e \subseteq t$' by ' $e \prec n$'. This gives the DRS in (1.7).

(1.17)
$$\begin{array}{|c|}\hline t \quad e \quad j \\ \hline t \prec n \quad e \subseteq t \quad \text{Joseph'}(j) \\ e\text{: turn-around'}(j) \\ \hline \end{array}$$

• We now turn to the construction of the DRS for the second sentence of (1.6.c). Its syntactic structure is shown in (1.18).

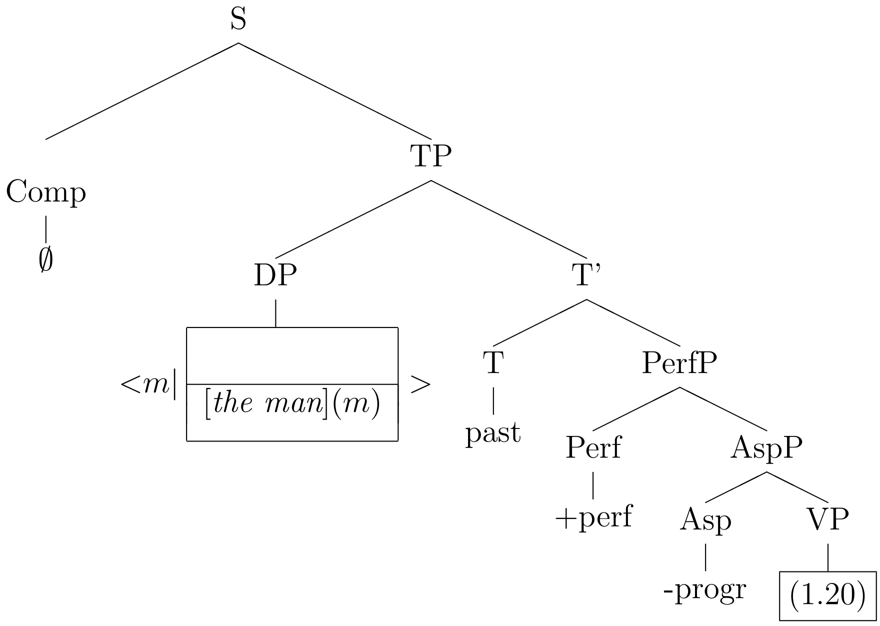
(1.18)



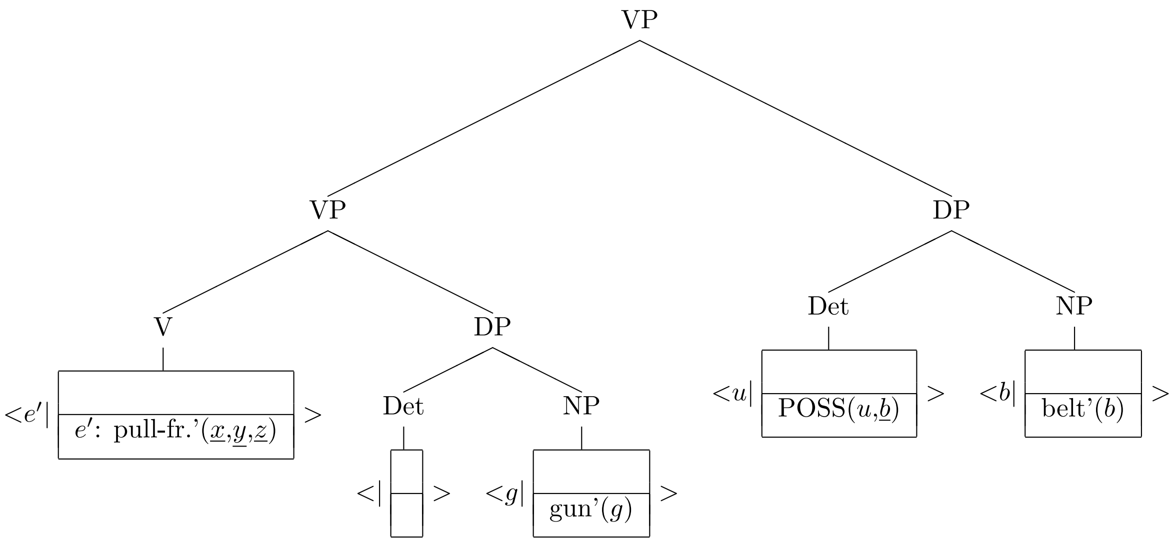
Lexical insertion, together with the 'pseudo-lexical' insertion for our stopgap treatment of the definite description *the man*, turns (1.18) into the tree below. We split the tree into two parts, (1.19) and (1.20), for space reasons.

Also for space reasons I omit from here on the subscripts $_{ref}$ from the occurrences of referential arguments in stores. (As mentioned earlier, the role that these subscripts play in stating the operations of the DRS Construction Algorithm isn't spelled out in this brief itroduction, so the subscripts may as well be suppressed.)

(1.19)

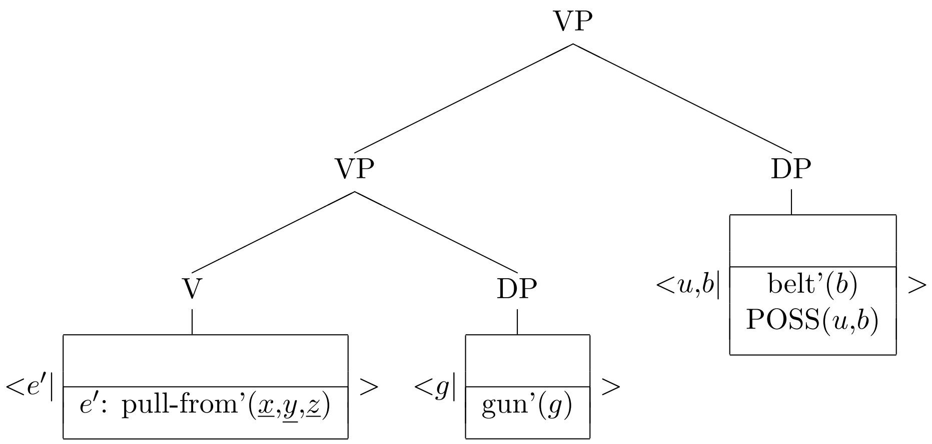


(1.20)



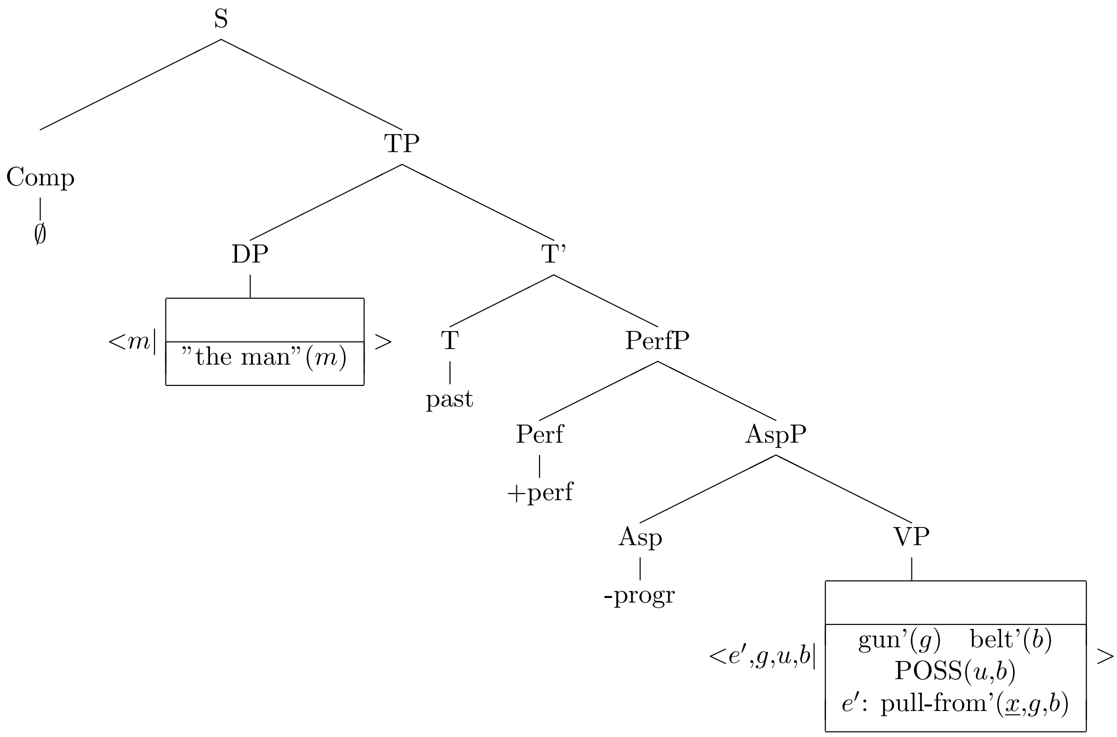
Combining the determiner *a* and the NP *gun* involves unification of the two referential arguments a and g. Combining *his* with the NP *belt* involves argument insertion of the referential argument b for the slot $\underline{b}$. These operations turn (1.20) into (1.21).

(1.21)



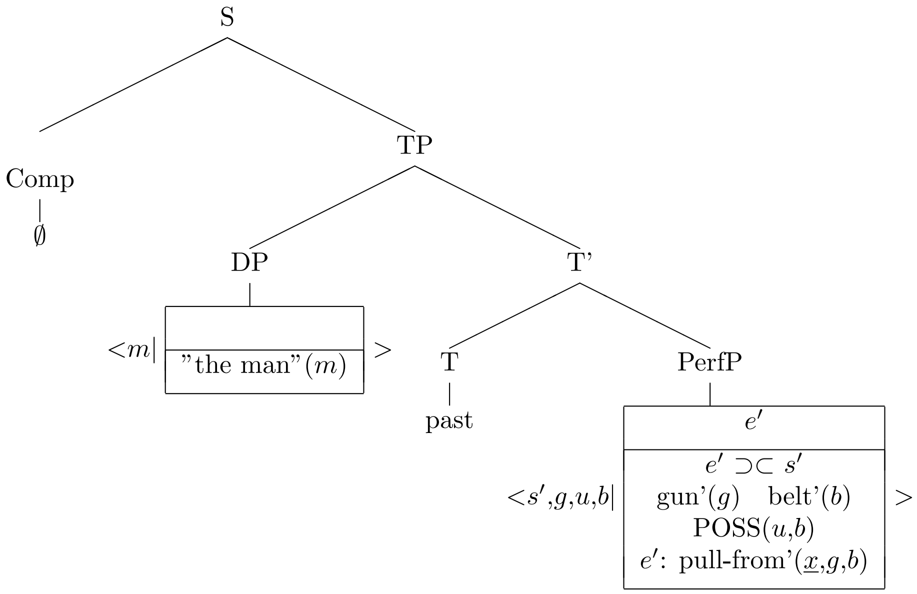
Combining the V representation of (1.21) first with the lower DP (the direct object) and then with the higher DP (the prepositional object) gives (1.22). (We now show the entire tree again.)

(1.22)



The feature -prog once again passes the VP representation up unchanged. +perf then changes the AspP representation to a result state representation. The referential argument of this representation is a state s', to the effect that an event e' of the kind described by the AspP representation has occurred. The fact that the result state starts the moment e' is complete is expressed by the relation symbol '$\supset\subset$' for 'abutment'. That is, the PerfP representation is as in (1.23).

(1.23)



This time the feature past acts in both ways that tense features typically do: (i) locating the eventuality in question in the past of the relevant time – here again the utterance time n – and (ii) linking it anaphorically to some time from the discourse context (its anaphoric link). In this case the eventuality in question is the state s'. Here the antecedent is the event e from the DRS for the first sentence. So we get as T' representation the one shown in (1.24).

$$
(1.24) \langle t', s', g, b, u \,|\,
\begin{array}{|c|}
\hline
e' \\
\hline
t' \subseteq s' \quad t' \prec n \quad e \subseteq s' \\
e' \supset\subset s' \\
\text{gun'}(g) \quad \text{belt'}(b) \\
\text{POSS}(u,b) \\
\\
e'\text{: pull-from'}(\underline{x},g,b) \\
\hline
\end{array}
\rangle
$$

Argument insertion for the subject DP, existential binding of the drefs in the store and elimination of t' transform (1.24) into the DRS (1.25). At this point it becomes possible to interpret the pronoun *his* as anaphoric to the subject phrase *the man*. In DRT this is implemented by setting the drefs for pronoun and antecedent equal to each other, by way of the Condition '$u = m$'.[9] This Condition can then be eliminated while occurrences of u are replaced by m.This gives the DRS in (1.25). Merging this DRS with the discourse context (the DRS for the first sentence) gives the DRS (1.26) for the first two sentences of (1.6.c).

$$(1.25)\quad \begin{array}{|c|} \hline t' \quad s' \quad g \quad b \quad m \quad e' \\ \hline t' \prec n \quad t' \subseteq s' \quad e \subseteq s' \quad e' \supset\subset s' \\ [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \quad \text{POSS}(m,b) \\ e'\text{: pull-from'}(m,g,b) \\ \hline \end{array}$$

$$(1.26)\quad \begin{array}{|c|} \hline t \quad e \quad j \quad t' \quad s' \quad g \quad b \quad m \quad e' \\ \hline t \prec n \quad e \subseteq t \quad \text{Joseph'}(j) \quad t' \prec n \quad t' \subseteq s' \quad e \subseteq s' \\ e' \supset\subset s' \\ [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \quad \text{POSS}(m,b) \\ e\text{: turn-around'}(j) \\ e'\text{: pull-from'}(m,g,b) \\ \hline \end{array}$$

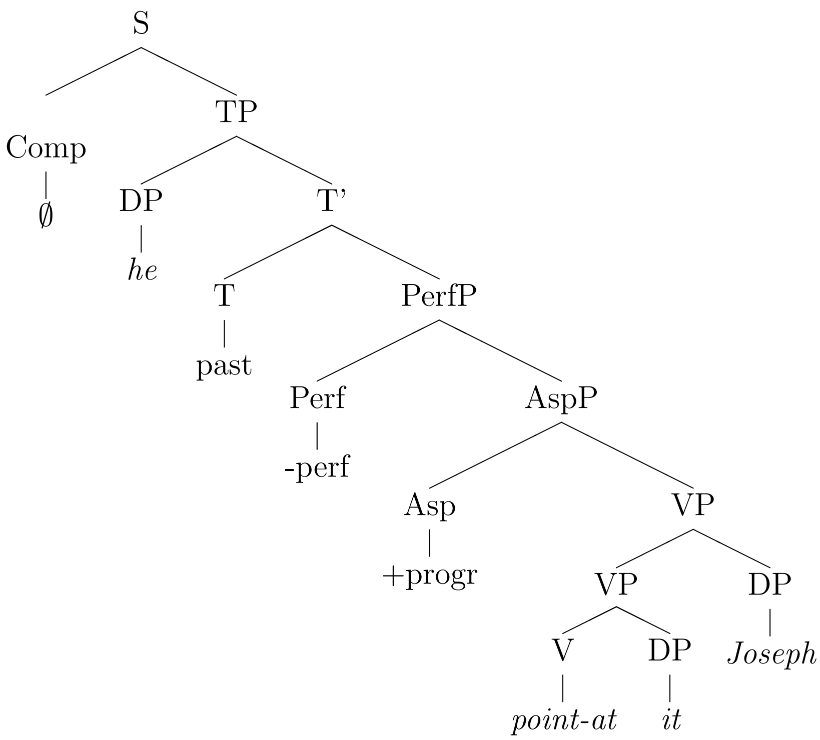
Note that this last DRS is nearly the same as the one we set as our target in (1.9). We get a DRS nearly identical to (1.9) when we eliminate the time drefs from (1.26), replacing their occurrences in DRS Conditions by the drefs for the eventualities whose location times they represent. The only difference remaining is then that in the resulting DRS in (1.27) the Condition '$e' \prec e$' of (1.9) is replaced in (1.27) by the pair of Conditions ' $e' \supset\subset s'$' and '$e \subseteq s'$' which says that e is located within the result state of e'. But that pair of Conditions clearly entails that e' temporally precedes e.

$$(1.27)\quad \begin{array}{|c|} \hline e \quad j \quad e' \quad s' \quad m \quad g \quad b \quad u \\ \hline e \prec n \quad e' \prec n \quad e' \supset\subset s' \quad e \subseteq s' \\ \text{Joseph'}(j) \quad [\textit{the man}](m) \quad u = m \\ \text{gun'}(g) \quad \text{belt'}(p) \quad \text{POSS}(u,b) \\ e\text{: turn-around'}(j) \\ e'\text{: pull-from'}(u,g,b) \\ \hline \end{array}$$

• We now come to the third sentence of (1.6.c). (1.28) gives its syntactic structure.

[9]The present version of DRT, which will get extended to MSDRT in the next chapters, follow the proposal of (Van Der Sandt 1992) to treat anaphoric expressions as presupposition triggers. Here I am skipping over a proper presuppositional treatment of the pronoun *his*. For more on this see the remarks below (1.30).

(1.28)



The construction of the upper VP representation is just as for the second sentence. The result is in (1.29.a). The +prog feature transforms this into the description of a state to the effect that a process is going on that will or could result in an event described by this VP representation. The result is in (1.29.b).

(1.29)a. $$<e'',v,j'|\begin{array}{|l|}\hline \\ \hline v = g \quad \text{Joseph'}(j') \\ e''\text{: point-at'}(\underline{x},v,j') \\ \hline \end{array}>$$

b. $$<s,v,j'|\begin{array}{|c|}\hline \\ \hline v = g \quad \text{Joseph'}(j') \\ s\text{: PROG}(\hat{}e.\begin{array}{|c|}\hline e \\ \hline e\text{: point-at'}(\underline{x},v,j') \\ \hline \end{array}) \\ \hline \end{array}>$$

The feature -perf passes the AspP representation in (1.29b) unchanged to the PerfP node. The past tense feature of (1.28), the syntactic structure for the 3rd sentence, then links this PerfP representation to an element from the discourse context. Here this is once more the event e from the first sentence. The final DRS for the 3rd sentence is shown in (1.30). Merge of this DRS with the DRS for the first two sentences completes the construction.

$$
(1.30)\quad \begin{array}{|c|}\hline s \quad z \quad v \quad j' \\ \hline s \prec n \quad e \subseteq s \\ z = m \quad v = g \quad \text{Joseph'}(j') \\ s\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: point-at'}(z,v,j') \\ \hline \end{array}) \\ \hline \end{array}
$$

• NOTE WELL. Among the construction steps that we have not motivated properly there are those that establish the links between anaphoric sentence constituents and elements from the discourse context. Why is it that the anaphoric constituents get linked with those elements from the discourse context and not with some others? The algorithm for constructing DRS from syntactic structures should give an answer to this question too.

Another way of putting this question: anaphoric constituents can be treated as *presupposition triggers* (Van Der Sandt 1992). Resolving the presuppositions they trigger consists in finding anaphoric antecedents for them. But what are the principles according to which these presuppositions are to be resolved?

The treatment of presuppositions proposed in (Van Der Sandt 1992) distinguishes two stages in the processing of linguistically triggered presuppositions, (i) the *representation construction* of these presuppositions and (ii) the *resolution* of the constructed presupposition representations, on the basis of their representations.

The result of the first stage is a so-called *preliminary* semantic representation, a representation in which all linguistically triggered presuppositions are represented. Only after the second stage has been completed – all represented presuppositions have been resolved – do we get, as final representation, a DRS, like the ones we ended up with above.

We will sometimes have to deal with presuppositions too. At that point we will introduce the relevant notation for presupposition representation and the mechanisms for presupposition resolution.

1.2.3 Back to Partee's marble example

Now that we have some idea of what DRSs are like, we briefly return to Partee's contrasting sentence pairs:

(1.3) a. Exactly one of the ten marbles is not in the bag. It is under the sofa.
b. Exactly nine of the ten marbles are in the bag. It is under the sofa.

Here we only present the completed sentence DRSs. (For a detailed discussion of the construction of these DRSs see (Kamp 2021*a*).)

We also ignore the tenses and eventuality-related aspects of these representations – a simplification that is not really problematic so long as we are prepared to make the plausible assumption that the present tenses in (1.3.a,b) all convey that the different predications expressed by the two discourses all hold at the utterance time.

(Readers may want see this as a multiple challenge: (i) Think of what these representations should be like when the tenses in (1.3.a,b) are to be dealt with along the lines adopted in our treatment of (1.6). But note well: A challenge not yet encountered here that is presented by he first sentences of (1.3.a) and (1.3.b) are the partitive noun phrases *exactly one of the ten marbles* and *exactly nine of the ten marbles*.)

The DRS for the first sentence is shown in (1.31).

$$
(1.31)\quad
\begin{array}{|c|}
\hline
x \quad Y \quad z \\
\hline
x \in Y \quad |Y| = 10 \quad \text{marble'}(Y) \quad \text{bag'}(z) \\
\neg \begin{array}{|c|}
\hline
\\
\hline
\text{in'}(x, z) \\
\hline
\end{array} \\
\hline
\end{array}
$$

Comments:

(i) Lower case italics are used for drefs that represent individuals. Upper case italics are used for drefs representing sets of individuals of more than one element.

(ii) the vertical slashes to the left and right of Y denote the cardinality of the set Y. (So '$|Y| = 10$' means that the set Y has 10 members.)

The reason why *it* in the second sentence of (1.3.a) can be interpreted as anaphoric to *one of the ten marbles* is that the referential argument for the noun phrase *one of the ten marbles* – the dref x in (1.31) – occurs in the Universe of the DRS (1.31). Pronouns in subsequent sentences have access only to antecedents whose referential arguments occur in the main Universe of the discourse context.

There is no dref for the missing marble in the DRS for the first sentence of (1.3.b), shown in (1.32). So *it* in the second sentence cannot be interpreted as referring to the missing marble in this case.

(1.32)
$$\begin{array}{|c|} \hline X \quad Y \quad z \\ \hline |X| = 9 \quad X \subset Y \quad |Y| = 10 \quad \text{marble'}(Y) \quad \text{bag'}(z) \\ \begin{array}{|c|}\hline \\ \hline x \in X \\ \hline\end{array} \ \Diamond_{\forall x} \ \begin{array}{|c|}\hline \\ \hline \text{in'}(x,z) \\ \hline\end{array} \\ \hline \end{array}$$

Comment: The Condition at the bottom of (1.32) expresses universal quantification over the members of the set X. This is an example of a *complex* DRS Condition. Conditions like this one are called *Duplex Conditions*. They consist of two DRSs, the *restrictor* DRS and the *nuclear scope* DRS, connected by the diamond in the middle, which represents the quantifier, consisting of the quantifying operator, here $\forall$, and the dref it binds, here x.

Another example of a Complex DRS Condition is the *Negation Condition*

$\neg \begin{array}{|c|}\hline \\ \hline \text{in'}(x, z) \\ \hline\end{array}$ in (1.31). Complex Conditions, like this one and the Universal Quantifier Condition of (1.32), are needed to give DRS languages the expressive power of the full First Order Predicate Calculus. (Without them all that could be expressed are existential quantifications over conjunctions of atomic predications.)[10]

The point of (1.3) is that the first sentences of (1.3.a) and (1.3.b) are logically equivalent in the sense that each entail the other, but that that doesn't make them equivalent for purposes of pronominal anaphora. It should be clear that the DRS in (1.32) confirms the entailment from the first sentence of (1.3.b) to the first sentence of (1.3.a); with a minimal amount of set theory we can derive from (1.32) that there is a unique marble from the set of ten that is not in the bag. But that isn't good enough for resolution of the pronoun.

The conclusion we can draw from this is that DRSs capture properties of discourse structure that go beyond the truth conditions they determine, that are essential for certain aspects of discourse interpretation.

1.2.4 Psychological Reality of DRT

The success that DRT has had in dealing with the phenomena we have rehearsed naturally gives rise to questions about cognitive significance. Human speakers must be able to record for themselves what entities have been introduced in the part of the discourse or text they have heard or read; otherwise they wouldn't be able to judge discourses like Partee's marble examples the way they do. Likewise, they must have a way of recording how eventualities are described, to be able to appreciate the difference between the Passé Simple and Imparfait sentences in (1.2) (about Jean opening his eyes) and to share the intuitions about the punctuality of Passé Simple events. These are aspects of our knowledge of language that DRT seemed to capture adequately and that for instance the standard applications of Montague Grammar cannot.

But does this tell us anything about the 'psychological reality' of DRT? Does it justify the assumption that human interpretations of sentences and texts take the form of DRSs and interpretation itself the form of DRS construction, along the lines of the few examples we have looked at here? This has been a controversy from the time DRT got its first formulation ((Kamp 1981*b*),

[10]It is peculiarity of DRS languages that Negation Conditions like the one above, in which the negation sign ¬ can be applied to any DRS, suffice for this purpose. For discussion see (Kamp & Reyle 1993), Ch.2.

(Kamp 1981*a*)). The debate over this question there has been over this question has been hampered by the problem that it is difficult to make precise exactly what the questions are. There was then, and still is now, so very little we understand about how information that we express in language is stored and manipulated in the human mind, and so long as we do not know more about this in general, most questions about psychological reality of structures or procedures posited by theories that deal with certain aspects of human mental behavior cannot be formulated precisely enough.

Only in some cases is it possible to refute cognitive significance in spite of our limited understanding of what the question should precisely come to. It is clear, for instance, that the construction algorithms of DRT cannot match step-by-step what goes in on in the minds of human interpreters. Human interpretation proceeds incrementally even within a single sentence. That is, if we assume as DRT does that interpretation involves both constructing a syntactic tree for the sentence and a semantic representation of its content which is related to this tree through compositional assignment, then both syntactic tree and content representation are built 'on-line' as the interpreter traverses the sentence presented to him 'from left to right'. What human interpreters quite clearly do not do is to *first* build a complete syntactic tree and *then* construct a semantic representation compositionally for this tree.

This need not mean, of course, that the compositional relation between syntax and semantics isn't part of our knowledge of language. Incremental syntactic and semantic parsing may be heuristic procedures for building pairs of syntactic trees and semantic representations which stand in this relationship; and part of the interpretation process may be the verification that the incrementally built pair is in accordance with the grammar – assuming that interpretation does take the form of constructing semantic representations that are compositionally derivable from syntactic trees.

But is this last assumption right? Are DRSs the semantic representations that human interpreters build? We are still at a loss how to ask that question in a sufficiently precise way.

The original motivation behind DRT was to develop an approach to natural language semantics that does greater justice to what goes on in the minds of human language users than Montague Grammar, where the official philosophy (and for all we can tell, also Montague's own deep conviction) that formal accounts of the syntax and semantics of natural languages should make no claims about the psychology of language users. The methodological reflections we have just engaged in indicate how difficult it is to argue convincingly that DRT has been successful in this sense. But that is compatible with the view that DRT has captured some aspects of human language interpretation that earlier approaches within Formal Semantics cannot do justice to.

The motivations for MSDRT, the framework to which this text is devoted, are superficially like the original motivations for DRT; in both case the aim was to capture cognitive aspects of language use that had not so far been charted in formal semantics. But the resemblance is no more than superficial. The mental structures that are modeled in MSDRT focus on the representation, organization and manipulation of propositional attitudes and entity representations ('file cards' in the terminology of Heim, Perry, Récanati and others) as constituents of the mental states of human agents, with a special concern for the implications of its hypotheses for the interpretation and production of language. In the mental states proposed by MSDRT DRSs are assumed as content specifications for the propositional attitudes of those states. Whether these DRSs should be seen as doing more than just capturing the truth-conditional contents of those attitudes, and establishing certain connections between the contents of different attitudes, is a question that will remain with us; we will have no better means to answer, or even formulate them than we have in the original DRT context. But for the aims of MSDRT this question does not matter all that much.

1.3 Model Theory

We have noted repeatedly that DRT is a 'Logical Form framework' for doing natural language semantics. In a Logical Form framework semantic representations, or Logical Forms as we also call them, are assigned to syntactic parses of sentences and a model-theoretic semantics is provided for the Logical Forms.

Central to each application of DRT to the semantics of some natural language fragment L is the choice and definition of a Logical Form Language LFL such that the Logical Forms assigned to well-formed expressions of L are well-formed expressions of LFL.

An application thus involves three separate tasks: (i) defining the syntax of L, (ii) defining the syntax and semantics of LFL and (iii) specifying a Construction Algorithm for converting sentences and discourses of L into LFs. In what we have said about DRT so far, none of these tasks were actually properly carried out. We have seen some sentences and multi-sentence discourses and for some of the sentences LFs were given. Also DRSs were shown for some sentences and discourses and in one case a detailed presentation was given of the DRS construction for a three sentences long discourse. But in all cases the fragment L and the Logical Form language LFL were left unspecified; the implicit assumption was that our sample sentences were well-formed expressions of L, and the DRSs well-formed expressions of LFL. And nothing whatever has been said so far about the model theory for the implied Logical Form Languages LFL.

In what follows we will continue our handwaving policy about L, assuming that if anyone insisted, a natural fragment could be specified which would include the sentences and texts that are being considered. But we will be less cavalier about the specification of viable LFLs, both with regard to their syntax and their model-theoretic semantics. Of particular importance will be the model theory.

A model theory for a syntactically defined language L involves two parts:

(i) specification of its *Model Class* $\mathcal{C}_{\mathcal{M}}$, and

a definition of the values of well-formed expressions of L in models from $\mathcal{C}_{\mathcal{M}}$. Our focus here will be on the structure of the models. But before we can start with that, we must first give a brief review of the main features of the semantic value definitions for which we need them. Here it is.

1. All LFLs that will play a role in the following are extensions of the First Order Predicate Calculus. This entails that some of their well-formed expressions are *formulas* – expressions that are assigned truth values in models. More specifically, we will assume that our semantics will always be two-valued, with as only truth values 1 (True) and 0 (False).

2. In addition, First Order Predicate Logic also has expressions other than formulas, such as individual constants and variables and predicates of various numbers of argument places (or 'arities'). For expressions of those other types the models will have to make available suitable values as well. On the whole DRS-languages have been fairly parsimonious as regards their repertoires of expression types, as distinct from the Typed Lambda Calculus and extensions thereof, such as Montague's HOI (see Section 1.1). But some of DRT's LFLs have expressions of types not found in the First Order Predicate Calculus.

3. Formulas and well-formed expressions of other types can be either *open* or

closed. In a closed expression there are no 'free variables' (in DRT: no free drefs). An open expression has one or more free variables (drefs).

4. The semantic values of closed formulas in models are truth values; and closed expressions of other types determine other kinds of values, appropriate to expressions of their type. But open expressions do not determine such values by themselves. They only determine values upon *assignments* of suitable values from the model to their free variables/drefs.

With this we turn to the models

Extensional models for Predicate Logic

The *extensional models* M for a language PL of First Order Predicate Logic are quite simple: they are structures of the form $<U,I>$, where

(i) U, the *Universe* of M, is some non-empty set and

(ii) I is a function that assigns to each non-logical constant Γ of PL an object of the right type – e.g. when Γ is an n-place predicate, then $I(\Gamma)$ is a set of n-tuples of elements from U; and when Γ is an individual constant, then $I(\Gamma)$ is an element of U.

The semantic value definition then takes the familiar form of the *truth definition* for First Order Predicate Logic, with their specific clauses for the logical constants ¬, &, ∀, .. such as:

P(x_1,...,x_n) is true in M under the assignment f iff $<f(x_1),...,f(x_n)> \in I$(P), for any n-place predicate P;

$\neg\phi$ is true in M under the assignment f iff ϕ is not true in M under f;

ϕ & ψ is true in M under the assignment f iff ϕ is true in M under f and ψ is true in M under f;

$(\forall x)\phi$ is true in M under the assignment f iff for all d $\in U_M$ ϕ is true in M under the assignment f[d/x],

(and so on).

Since something was said in Section 1.1 about the types of Typed Lambda Calculus, let us add a word about the extensional models for languages of this system too. In Section 1.1 we reviewed the type systems of the Extensional Typed Lambda Calculus and of Montague's intensional system HOIL. We noted that the types of these systems determine Domains, with entities whose structure reflects the structure of the canonical denotations of the types (such as ‘<e,t>’, ‘<<e,t>,t>’ and so on). But nothing more was said about those Domains. We can now be more explicit. Each extensional model M for a language of the Extensional Typed Lambda Calculus has a Domain $D_{M,\alpha}$ for each type α of the Extensional Calculus, and these Domains must be related to each other in the manner indicated in Section 1.1 ($D_{M,<\alpha,\beta>}$ = the set of all functions from $D_{M,\alpha}$ into $D_{M,\beta}$).[11]

1.3.1 History models for Tense Logic and for DRT

• The extensional models for languages of First Order Predicate Logic can be thought of as describing single situations, about which their formulas can be either true or false, like snapshots of the world as it is or could be at a single moment. For the DRS languages of DRT and MSDRT that we will be using as Logical Form languages, models cannot be this simple.

The principal reason for this is that formulas of these languages (their DRSs) must be able to talk about more than one time at once. Natural language sentences and bits of discourse often do this, for instance when they contain more than one clause and where the verbs in these clauses have different tenses (recall the examples in (1.6)). The question whether such a sentence is true when uttered at some given time t will in general depend on what the world is like at other times than t. The models of a model theory that can do justice to this must be able to represent not just single ‘time slices of worlds’, but temporally ordered sequences of such slices.

Such models, which represent *world histories* rather than single snapshots of histories, will be called *history models.*

The first history models arose in connection with *Tense Logic* (nowadays also: *Temporal Logic*) and we will discuss those first. The history models for our DRS languages are more complex than these and we will turn to those next. Two issues that will be important in connection with the latter:

[11] We will not have anything to say directly here about models for Montague's HOIL. For HOIL the story is more complex. Extensional models for HOIL are closely related to *intensional* models for the Extensional Typed Lambda Calculus, to which we will, step by step, be making our way.

1. the Continuity Principle
2. the difference between verbal and non-verbal predicates. Detemporalization of non-verbal predications.

Tense Logics vary in form, but the general model-theoretic issues that matter here are shared by all of them.

By way of illustration, consider what is perhaps the best known system of Tense Logic, and at the same time also the simplest one: Arthur Prior's 'P,F-Calculus'. This is an extension of the classical propositional calculus with two 1-place 'tense operators' P and F.

Syntactically P and F work like negation. They modify formulas: if ϕ is a formula, then so are Pϕ and Fϕ.

The intuitive meaning of Pϕ is 'it was the case that ϕ' and that of Fϕ 'it will be the case that ϕ'. Put more formally: Pϕ/Fϕ is true at a time t iff there is a time t' before/after t at which ϕ is true.

A model for the P,F-Calculus provides information about what the world is like at different times in a very simple way. It represents time as a linearly ordered structure' $<$T,$\prec$$>$, where T is the set of 'temporal instants' and $\prec$ the earlier-later relation between them. And for each time t in T they specify for each propositional constant q from the set Q of propositional constants of the language whether q is true or false at that time. That is: for each t M provides an 'extensional model' M_t for the propositional calculus, a function from Q into the set $\{0,1\}$ of the two truth values true (1) and false (0).

To repeat: a model M for the P,F-Calculus consists of (i) an instant structure $<$T,$\prec$$>>$ and (ii) a function that maps each $t \in$ T to a function M_t from the set Q of propositional constants of the language into the set $\{0,1\}$ of the two truth values.

Similar considerations also apply to a wide range of other Logical Form languages in which temporal relations can be expressed. Among them are *Priorean Tense Predicate Logics*, which extend classical Predicate Logic with P and F, or with some other set of tense operators.

Here a model M is a function from the set T of an instant structure $<$T,$\prec$$>$ to extensional models M_t for the Predicate Calculus.

Models for languages of Tense Logic – families of extensional models indexed by the instants of some linear time structure $<\mathrm{T},\prec>$ – are called *history models*. Formally a history model will be a pair $<<\mathrm{T},\prec>,\{M_t\colon t \in \mathrm{T}\}>$, where for each $t \in \mathrm{T}$ M_t is an extensional model for the given language.

In a history model M it is possible to evaluate formulas of the language for truth or falsity at any instant t of M's time structure. Intuitively, evaluating ϕ at t should be thought of as evaluating an utterance of ϕ at t, or as evaluating ϕ on the assumption of its being uttered at t.

Here are a couple of examples of tense-logical formulas and the semantics imposed on them by its model theory: (i) let M be a history model for Prior's P,F-Calculus and consider the formula $\neg\mathrm{P}\neg\mathrm{P}q$. It is not difficult to see that this formula is true at t in M iff for every t' such that $t' \prec t$ there is a t'' such that $t'' \prec t'$ and q is true in M at t''. (ii) Consider the formula $\mathrm{P}\neg\mathrm{F}\neg q$. This formula is true in M at t iff for some time $t' \prec t$ it is the case that for all $t'' \succ t'$ q is true at t'' in M. Note that $\mathrm{P}\neg\mathrm{F}\neg q$ entails q.

For our DRS languages we also need models involving time structures $<\mathrm{T},\prec>$. But the details are somewhat different. This has to do with the fact that the *ontology* of these languages is different. Recall that our DRSs have drefs for events and states, as well as for 'individuals' in general. In this respect they are languages of *many-sorted Predicate Logic*, whose models have Universes of individuals that are subdivided into the different ontological sorts. Each such Universe U has a subset consisting of events and a subset consisting of its states.

As regards the way they handle events and states, the models for these languages are like models for many-sorted Predicate Logic. But they differ from such models in the way they handle time. DRS languages have drefs for times (representing intervals of time rather than single instants), which means that the ontology of their models must include times also among their sorts. But in addition times must also be available as indices, at which DRSs can be evaluated for truth or falsity, just as for the languages of Tense Logic. In other words, the models for these languages must be history models of the form $<<\mathrm{T},\prec>,\{M_t\colon t \in \mathrm{T}\}>$, in which the Universes of extensional models M_t include $<\mathrm{T},\prec>$ as one of their ontological sorts. Time is thus part of these models 'twice over', (i) as index set and (ii) as one of the sorts included in their Universes, to which it is possible to refer in the language that the models are for.[12]

History models are instant-indexed sets of extensional models. All those models must provide at least the information needed in the truth definitions for languages of First Order Predicate Logic, and in the same form as the extensional models for those languages. This means in particular that each such model has a Universe (of type e), subdivided into different ontological sorts, among which at least the sort of times and that of eventualities (in its turn subdivided into events and states).

In general the Universes U_{M_t} of the different indexed extensional models M_t belonging to a history model M may vary. An entity **d**[13] that belongs to the Universe U_{M_t} of model M_t need not belong to the Universe $U_{M_{t'}}$ for some other time t' of the time structure of M, because **d** doesn't exist at time t' (according to what M says). But if that is what belonging to U_{M_t} means, then there are a number of questions that have to be addressed, which have to do with existence in and through time.

One of these questions has to do with 'temporal continuity'. Let EX$(\mathbf{d})_M$ be the set of times of the historical model M at which **d** exists: EX$(\mathbf{d})_M$ = {t $\in$ T: $\mathbf{d} \in U_{M_t}$}. What kind of' temporal profiles' can such sets be expected to have? What are the possible distributions of such profiles over the time structure $<$T,$\prec$$>$ of M?

One answer to this question is given by the *Continuity Principle*. According to the Continuity Principle, EX$(\mathbf{d})_M$ must be an interval of the time structure $<$T,$\prec$$>$ of M:

(1.33)(Continuity Principle)

When an entity **d** (belonging to such a sort) belongs to the Universes

[12]The double presence of time in these history models is further complicated by the fact that the members of the index set T are assumed to be instants of time, whereas the times that are referred to by the time drefs of the DRS languages are in general temporal intervals, which contain instants as their members. This further complication can be dealt with straighforwardly in actual applications of the framework and need not detain us in the present discussion.

[13] Since italic letters and letter combinations are used for discourse referents, we will often use bold face letters to denote elements of models or of the world that a sentence or discourse is about, so as to set these clearly apart from the discourse referents (of which they will often be the values under assignments or embedding functions). We will not be 100 oercent consistent with this practice, but bold face will always entities that are talked about, and never drefs.

of extensional models M_t and $M_{t'}$ and $t \prec t'' \prec t'$, then $\mathbf{d}$ belongs to the Universe of $M_{t''}$

How plausible is the Continuity Principle? The answer to this question, or at least its justification, varies from one ontological sort to another. For physical objects the principle is widely accepted as true (mostly tacitly), although even here it hasn't gone unchallenged. But what is plausible for physical objects need not be so for other categories. Events and states are notorious in this connection. For events the Continuity Principle has often been assumed as well. (Recall the Russell-Wiener method discussed in section 1.2 for the construction of instant structures from event structures, which assumes that the principle holds for events.) But even if it is not unreasonable to assume that the Continuity Principle applies to events, it isn't obvious either. Take for instance an event of your writing something. You started yesterday and sat down half an hour ago to complete your piece, after a good night's sleep. Is your writing yesterday, which went on until 10.30, and your writing today, which started at eight this morning, one single event, which will go on until I am done, and which was 'dormant' from when I went to bed till when I sat down in front of my laptop again at eight? Or is this a case of two events, one ending when you stopped writing last night and the other one starting this morning, both in the service of getting the piece done, but nonetheless still distinct events? In my experience, the one-event story is usually the more plausible one; but there are also many cases where it is hard to tell which story is better.

For yet other types of entities, such as states of various kinds, including mental states like intentions and beliefs, the Continuity Principle will raise further issues. But there too it has proved helpful as a guiding principle, which usefully structures our conceptions of the structure of those various categories. In the discussions of MSDRT in the coming chapters I will treat it as a default. In fact, we will treat the Continuity Principle as a general default principle, which applies to all ontological sorts for which it makes sense to ask when in time members of them did or did not exist. If there are exceptions to the principle, then these will always have to be mentioned explicitly and argued convincingly. But in this MSDRT introduction there will be no cases of this.

One important consequence of the Continuity Principle is that eventualities have *durations*: The set $\mathrm{EX}(\mathbf{ev})_M$, where $\mathbf{ev}$ is either an event or a state of

M, is always an interval of the time structure[14]. We denote the duration of an eventuality **ev** in M as 'dur$(\mathbf{ev})_M$'; often the subscript $_M$ is omitted.[15]

1.3.2 Truth Definitions for DRS Languages

• Truth definitions for DRSs languages closely follow the definition of truth-relative-to-an-assignment for languages of Predicate Logic. But one difference is that truth definitions for DRS languages work with partial assignments rather than total ones. These partial assignments are restricted to drefs that occur in the DRS that is being evaluated, and is therefore in all cases of practical interest finite.

Connected with this use of partial assignments is a conceptual difference. The assignments of DRT can be seen as *embeddings* of DRSs within models, as if the DRS was a picture of some part of the world represented by the model, and the embedding is a way of overlaying that part of the world with the DRS; and when the overlay is a perfect match, then this verifies the DRS as true of this world. We should not forget, however, that this 'picturesque' way of looking at the truth definitions for DRS languages may suggest a bigger difference with the familiar truth definition for Predicate Logic than is actually warranted. A more sober, but formally equivalent way of saying that an embedding f verifies the DRS K in the model M is that f is an the assignment relative to which K is true in M. To be noted in this connection is that the picture metaphor is apt only so long as the DRS does not contain an Complex DRS Conditions (such as negation Conditions or Universal Quantifier Conditions). The verification of such Conditions can be understood only in the more abstract sense exemplified by the truth definition for Predicate Logic.

But this comparison between the truth definitions for DRS languages and the truth definitions for languages of the Predicate Calculus only considers extensional models and DRSs that can be evaluated in such models. And for the DRSs of the DRS languages that will be considered in this MSDRT introduction that is hardly ever the case. Nearly all DRSs we will encounter have Conditions involving time drefs that do not represent the utterance time of the sentence or discourse from which the DRS has been derived as its Logical Form. And Conditions containing such drefs can only be evaluated in

[14] In the sense that if two instants t and t' belong to the interval and t'' is some other time of the time structure such that $t \prec t'' \prec t'$, then t'' belongs to the interval as well.

[15] In much of the formal semantics literature the symbol τ is used to denote the durations of eventualities.

history models, and not in the extensional models of which these are made up.

To see this more clearly, let us look more closely what kinds of Conditions occur in DRSs like those we constructed as Logical Forms for the discourses in (1.6). We can distinguish three types of atomic Conditions:

(i) Conditions that represent verbal predications from the represented sentence, (for instance 'e: turn-around'(j)', 'e': pull-from'(m,g,b)'

(ii) Conditions that represent non-verbal representations, such as 'Joseph'(j)', 'belt'(b)' or 'POSS(m,b)' from (1.26) (and expressed by the first possessive pronoun *his* in (1.6.c)) or the Condition 'in'(x, z)' from (1.31) (where *in* is used with the spatial sense according to which a marble can be in a bag); and

(iii) 'temporal order' and 'temporal location Conditions', such as '$t \prec n$', '$e \subseteq t$' or '$e' \supset\subset s'$' from (1.26).

In addition there are various types of complex Conditions. Some of these belong to DRS languages because they extend First Order Predicate Logic. These are verified in essentially the same way as in Predicate Logic – three of the satisfaction conditions for the corresponding logical constants of Predicate Logic were mentioned above under *Extensional models for Predicate Logic and the Lambda Calculus* in Section 1.3. As the verification conditions for those complex Conditions have been described in detail in some of the DRT literature (see in particular (Kamp & Reyle 1993), Ch .2), we skip those verification conditions here. Other complex Conditions that will play an important part later on are specific to MSDRT. Their verification conditions will be dealt with eventually.
Something, however, should be said here and now about the verification conditions of the three types of atomic Conditions. So let M = <<T,$\prec$>,{M_t: $t \in$ T}> be a history model and let t be an instant of T. We consider for each of the three types of atomic Conditions what it is for that Condition to be verified in M at t by an embedding f that assigns values in M to the drefs occurring in those Conditions.

We start with Conditions of type (iii). First consider a Condition of the form '$t' \prec t''$'. Let f assign appropriate values to the drefs t' and t''; that is, for these drefs, f assigns to each an interval from the time structure <T,$\prec$> of M. Let $f(t')$ = **t'** and $f(t'')$ = **t"**. Then:

(1.34) f verifies '$t' \prec t''$' in M at t iff **t'** $\prec$ **t"**.

Note that this verification condition does not involve t. In other words, embedding functions verify this Condition in a model M at one time t iff they verify it at any other time. In this sense the Condition is *eternal*. And it should be clear from what has been said so far that this is true also for Conditions that are like (1.34) except for the temporal relation. (It should be obvious what the verification conditions are when $\prec$ gets replaced by $\supset\subset$, $\subseteq$ and other such temporal relation terms, and that all the resulting Conditions are eternal.)

But this is not so for all Conditions that express temporal relations between time representing drefs. Consider '$t' \prec n$'. Here the verification conditions are those in

(1.35)f verifies '$t' \prec n$' in M at t iff $f(t') \prec f(n)$ and $f(n) = t$.

The verification conditions in (1.35) do depend on the evaluation time t, but that is because of the indexical status of the special utterance time representing dref n.

We now turn to the Conditions of type (i), those that represent predications expressed by verbs. Here it doesn't matter which example we take. I choose 'e': pull-from'(m,g,b)'. Intuitively it seems clear what the verification conditions of this Condition should be. 'pull-from' ' is a 4-place predicate with one event argument and three other arguments, to be filled by individuals of various kinds, but here we will not worry separately about the kinds of those. The interpretation function I of a model M for a DRS language containing this predicate will have to specify which tuples $<\mathbf{ev},\mathbf{x},\mathbf{y},\mathbf{z}>$ belong to the extension of 'pull-from' '. But the question is: Where in M should this information be stored? M is a time-indexed family of extensional models, each of which comes with a universe and an interpretation function. Which of these interpretation functions should specify this information?

There is no obvious answer to this question. Intuitively the 4-place relation expressed by 'pull-from' ' holds at times when **ev** exists. And since **ev** is by definition an event that stands in the 'pull-from' ' relation to the other three members of $<\mathbf{ev},\mathbf{x},\mathbf{y},\mathbf{z}>$, it seems reasonable to assume that the information that $<\mathbf{ev},\mathbf{x},\mathbf{y},\mathbf{z}>$ belongs to the extension of 'pull-from' ' should be part of the interpretation functions I_{M_t} of models M_t for which **ev** exists at t. That is, $<\mathbf{ev},\mathbf{x},\mathbf{y},\mathbf{z}>$ should belong to I_{M_t}(pull-from') iff t belongs to dur(**ev**). But note well that is a stipulative decision about how we want to set up our history models. And of course the same information gets multiply stored this way. If $<\mathbf{ev},\mathbf{x},\mathbf{y},\mathbf{z}>$ belongs to I_{M_t}(pull-from') means that **ev** is an event of **x** pulling **y** from **z** and that $t \in$ dur(**ev**) and this information is repeated for all times t belonging to the duration of **ev**.

Let us assume from now on that this is how the extension of verbal predicates is encoded in history models. Then the verification conditions for such predications can be stated as in (1.36) for those involving *pull-from'*.

(1.36) f verifies 'ev: pull-from'(x,y,z)' in M at t iff f assigns an event from M to ev and individuals from M to x,y and z and $<f(ev),f(x),f(y),f(z)> \in I_{M_t}$(pull-from').

But note that there is a certain slack here. We might as well have stated these verification conditions as:

f verifies ev: pull-from'(x,y,z)' in M at t iff f assigns an event from M to ev and individuals from M to x,y and z and there is some t' in T such that $<f(ev),f(x),f(y),f(z)> \in I_{M_{t'}}$, or as

f verifies ev: pull-from'(x,y,z)' in M at t iff f assigns an event from M to ev and individuals from M to x,y and z and for some/all t' in T such that $t' \in$ dur(ev), $<f(ev),f(x),f(y),f(z)> \in I_{M_{t'}}$.

Each of these options – (1.36) and the two alternatives just mentioned – is open to us. We will stick with (1.36).

Much the same applies to state predications, as distinct from event predications. Consider for instance the verb *to know* in its use of acquaintance between people, exemplified in the following variant of our 'Joseph' discourses in (1.6).

(1.37) Joseph turned around. The man had pulled his gun from his belt. Joseph knew him.

State verbs too are treated in DRT as predicates involving an eventuality argument, but with them the eventualities are states. So the satisfaction of *know* as it occurs in (1.37) involves triples $<s,x,y>$ consisting of a state and two individuals. The same question then arises where the information that such a triple belongs to the extension of *know* should be encoded in history models. And the answer to that question we adopt is the same as for event verbs.

With state predications there is however a complication that doesn't arise for event verbs. One difference between events and states as these categories are understood in DRT has to do with the difference between perfective and non-perfective aspect, as analyzed prominently in the work of Krifka (Krifka 1989), (Krifka 1998). One manifestation of this distinction in DRT is the difference between the temporal location Conditions for events and states – an event e is related to its location time t by the Condition '$e \subseteq t$', a state s by the Condition '$t \subseteq s$'. But there is also another difference between states and events, often referred to as that between homogeneity and non-homogeneity: An event e that is given by a description $\Phi(e)$ will have no proper temporal sub-events e' that can also be described by Φ (i.e. as $\Phi(e')$). That is the non-homogeneity of events. In contrast, states are homogeneous in the sense that when a state s is given by a description $\Phi(s)$, then it will have proper temporal sub-states s' that can also be described by Φ.

In DRT a strong form of the homogeneity of states has been adopted in the form of the Principle of the *Plenitude of States*. Plenitude of States is the following constraint on history models:

(1.38) Let M be a history model and t an instant from its time structure. Let P be an n+1-place state predicate of the DRS language for which M is a model and let the tuple $<\mathbf{x}_1, .., \mathbf{x}_n>$ belong to I_{M_t}(P). (So **s** is a state of P holding between the individuals $\mathbf{x}_1$, ..., $\mathbf{x}_n$.) Let **t'** be an interval of time from the time structure of M such that $\mathbf{t} \subseteq$ dur(**s**). Then there is a state **s'** such that dur(**s'**) = **t'** and the tuple $<\mathbf{s'}, \mathbf{x}_1 \ldots \mathbf{x}_n>$ also belongs to I_{M_t}(P).

Put in simple terms: For any state of description Φ and duration **t** and any time **t'** included in **t** there is a state of description Φ and duration **t'**.

Verification Conditions for Predications of Type (ii).

Lastly we turn to the third type of atomic Condition, that of the non-verbal predications. Here too it doesn't matter much which example we look at. Let's take the possession relation expressed earlier by the DRS Condition 'POSS(m,b)' from (1.26).

From the tense-logical perspective that we adopted above to motivate the introduction of history models this type may seem the most straightforward because it looks most directly in line with the traditional treatments of Tense Predicate Logic (Cocchiarella 1965): predications can hold at some times and not at others and that is captured by the interpretation functions I_{M_t} of the different extensional models that belong to the given history model M: the satisfying pair of individuals $<y,z>$ belongs to I_{M_t}(POSS) iff y and z stand/stood in the relation *POSS* at time t according to M. This suggests the verification conditions for such predications in (1.39):

(1.39) f verifies 'POSS(x,y)' in M at t iff f assigns individuals from M to x and y and $<f(y),f(z)>$ belongs to I_{M_t}(POSS).

But leaving things at this won't do, for the following reason. Recall the DRS in (1.25) for the second sentence of the 'Joseph' discourse (1.6.c), repeated here for easier reading.

$$(1.25)\quad \begin{array}{|c|} \hline t' \quad s' \quad g \quad b \quad m \quad e' \\ \hline t' \prec n \quad t' \subseteq s' \quad e \subseteq s' \quad e' \supset\subset s' \\ [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \\ \text{POSS}(m,g) \quad \text{POSS}(m,b) \\ e'\text{: pull-from'}(m,g,b) \\ \hline \end{array}$$

Suppose that the DRS (1.25) is the Logical Form of an utterance made at the time **t** of the time structure of a history model M and that we therefore want to evaluate this DRS in M at **t**. Part of this evaluation will involve the predication 'POSS(m,g)' in this DRS. But how should it be evaluated? Not at the utterance time **t**, for there is no reason that the predication holds at **t**. (For all we know, Joseph was the one to survive the encounter and the man has been dead ever since and doesn't possess anything any more, not even his gun.) Rather, the evaluation time for this predication should be at the time in the past of the utterance time **t** that is introduced by the past tense of the second sentence of (1.6.c).

Detemporalization

Here is the way we have adopted to deal with this problem. It involves three parts. The first concerns the form of non-verbal Conditions like 'POSS(x,y)'. How should this form be changed, so that it can indicate at what time or times they are supposed to hold? The second part is how these revised forms can express where in time the predication is supposed to hold. And the third part has to do with the interpretation of sentences and discourses. How does it determine where such predications are supposed to hold when they occur in the context of the sentences or discourses in which they occur?

First the new form for non-verbal predications.It consists in transforming Conditions of the form 'P'(x_1,...,x_n)', where P' is a non-verbal predicate and x_1,..., x_n are its arguments, into Conditions of the from 's: P'(x_1,...,x_n)', where s is a state dref. Intuitively, 's: P'(x_1,...,x_n)' means that s is a state to the effect that the predication 'P'(x_1,...,x_n)' is currently true. In other words, the predication is true throughout the duration of s.

Formally, the switch from 'P'(x_1,...,x_n)' to 's: P'(x_1,...,x_n)' is that from an n-place to an n+1-place relation, where the extra argument position of the new relation is restricted to states. This means that the DRS languages for dealing with natural language fragments containing non-verbal predicates – and that means in practice: any natural language fragment – now contain such n+1-place state predicates instead of the n-place predicates that predicate words other than verbs had been assumed up to now.

This change also affects the models. A history model for a DRS language of the new kind will need to have states to allow verification of the new state predications that get added in the transition to the new languages. We assume that these new models still contain interpretations for the original n-place predicates, as described in the paragraph above (1.40). These interpretations now serve as basis for the extensions of the new n+1 predicates. In other words, let M be such a model and suppose that **T'** is an interval from the time structure of M throughout which the tuple $<\mathbf{x}_1,..,\mathbf{x}_n>$ belongs to the extension of P'. (That is, $<\mathbf{x}_1,..,\mathbf{x}_n>$ belongs to $I_{M_\mathbf{t}}$ for every {bf t in **T'**.) Then M has a corresponding state $s_{P'}$ whose duration is **T'**. And the n+1-tuple $<s_{P'},\mathbf{x}_1,..,\mathbf{x}_n>$ belongs to $I_{M_\mathbf{t}}$(P') for each time **t** in **T'**.[16]

[16]Note that this specification non.verbal predication states satisfies the Plenitude of States Principle given in the treatment of state verbs we described above.

With the new notation the Logical Form of the second sentence of (1.6.c), the previous form (1.25) of which, we repeated a couple pages before, now takes the form shown in (1.40)[17]

$$(1.40)\quad \begin{array}{|c|} \hline t' \quad s' \quad s_g \quad s_b \quad g \quad b \quad m \quad e' \\ \hline t' \prec n \quad t' \subseteq s' \quad e \subseteq s' \quad e' \supset\subset s' \\ [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \\ s_g\text{: POSS}(m,g) \quad s_b\text{: POSS}(m,b) \\ e'\text{: pull-from'}(m,g,b) \\ \hline \end{array}$$

This is a step forward, but it still isn't the whole story. For (1.40) tells us nothing about where in time the states s_g and s_b are located. In the present instance it is intuitively clear that they hold at the time described by the discourse (1.6.c), which starts with the event e introduced by the first sentence. In the revised DRS for the first two sentences together, shown in (1.41) this can be made explicit by means of the Conditions '$e \subseteq s_g$' and '$e \subseteq s_b$'.

$$(1.41)\quad \begin{array}{|c|} \hline t \quad e \quad j \quad t' \quad s' \quad s_g \quad s_b \quad g \quad b \quad m \quad e' \\ \hline t \prec n \quad e \subseteq t \quad \text{Joseph'}(j) \quad t' \prec n \quad t' \subseteq s' \quad e \subseteq s' \\ e' \supset\subset s' \\ [\textit{the man}](m) \quad \text{gun'}(g) \quad \text{belt'}(b) \\ s_g\text{: POSS}(m,g) \quad s_b\text{: POSS}(m,b) \\ e \subseteq s_g \quad e \subseteq s_b \\ e\text{: turn-around'}(j) \\ e'\text{: pull-from'}(m,g,b) \\ \hline \end{array}$$

But what are the principles which determine the temporal location of non-verbal predications in discourse? That is a notorious problem that has had a fair bit of attention in the semantics literature (Enç 1981), (Musan 1999), (Tonhauser 2002). But an algorithm for determining temporal locations of nominal and other non-verbal predications in discourse, with a reasonably

[17](1.40) doesn't implement the new regime fully in that it still contains non-verbal predications that haven't been converted to the new form, among them 'gun'(g)' and 'belt'(b)'. For these predications, which involve the sortal nouns *gun* and *belt*, the transformation into state predications is less urgent, since for practical purposes sortal nouns are *eternal* predicates: if an entity satisfies such a predicate, it will do so throughout its existence. For predications involving eternal predicates it doesn't matter where they are evaluated, so long as their arguments exist at the evaluation time.

wide coverage, is to our knowledge still missing. We just note this problem for Logical Form construction, leaving it to be dealt with in some other place than here. In what follows we will handle the location of such predications ad hoc.[18]

1.3.3 Intensional History Models

This much about history models. But history models aren't the final word about the model theory we want for our representation languages. There are a number of different reasons for this. One is our conception of the world as open towards the future: What the future will be like, how it will unfold, is indeterminate in many ways, and we ourselves can help shape it through the actions we decide to perform. In this respect the future is for us fundamentally different from the past. The past is what it is, and nothing we do now can alter it. History models cannot do justice to this asymmetry. According to history models there is, at any time t of its time structure, just one future course of events, just as there is just one past. To capture the difference we need models that has, for any history and time t within it, a bundle of alternative ways in which that history could go on after t. One way to model this is to assume models that are sets of history models and where two history models can share a part – an initial segment from the beginning of time up to some time t – but then diverge after t. This is how we will model the distinction between closed past and open future.

A second consideration is that many constructions found in natural languages are *intensional*: they involve predicates and operators whose arguments are

[18] A fist rule is that non-verbal predications either hold at the location time of the eventuality introduced by the main verb of the clause in which the non-verbal predicate occurs or else at the utterance time. But this is no more than a first approximation. A flavor of the complexities that such an algorithm for determining the times of such predications will have to cope with is the following utterance (by, let's assume, a somewhat more than middle-aged gentleman)

(1.42) I first met my wife in Palo Alto

In our society spouses are normally met before they are wed. So the time at which the referent of *my wife* stood in the 'wife' relation to the speaker was presumably after the time of her first meeting with her later husband. But can we infer from this that the 'wife' relation must hold at the utterance time? No, not necessarily. It is quite possible that the speaker got a divorce in the mean time. Perhaps the speaker is talking about an earlier epoch of his life, when he and the referent of *my wife* were still married.
This example is still moderately simple. It is just the beginning of a much longer story.

entities that aren't part of history models. Prominent examples from the literature are modal operators like *it is necessary that* and the sentence adverb *necessarily*, counterfactuals and causality-related vocabulary. All these involve sentential or clausal complements whose semantic contributions cannot be explained in terms of history models (let alone single extensional models). These contributions are usually described as 'propositions' or 'propositional', but these descriptions aren't of much help unless they come with a precise definition of what propositions are. The intensional approach identifies propositions (and related entities such as properties and individual concepts) in terms of possible worlds. In particular, propositions are, according to the intensional perspective, functions from possible worlds to truth values (or, equivalently when sentences are assumed to be always either true or false, sets of possible worlds). The proposal is an old one, but it has been criticized more or less from the start. For one thing, it doesn't take much to see that it causes problems for the analysis of beliefs and other 'propositional attitudes' as relations between attitude holders and what is expressed by the complements of *believe* and other attitudinal verbs in attitude attributing sentences. If we assume that the clausal complements of these verbs contribute the 'propositions' they express to the truth conditions of attitude attributions and these propositions are defined as sets of possible worlds, then far too many beliefs (and other attitudes) are predicted to collapse into one. The most notorious problem are beliefs in mathematics, and closely related attitudes like conjectures; here the propositions involved are typically 'modally absolute', that is either true in all possible worlds or else false in all possible worlds. This leaves only two non-equivalent mathematical propositions, the necessarily true and the necessarily false one, with the implication that someone who makes one mathematical conjecture and rejects another, and where both conjectures are in fact true, is in a mental state that nobody can really be in.

This consequence for an account of belief in mathematics seems so patently absurd that one would have thought that no account that entails it could survive it. But not so. Intensional analyses of propositional attitudes and a host of related issues, where the identification of propositions with sets of possible worlds is similarly problematic have enjoyed an astounding popularity. On the one hand the identification of propositions with sets of possible worlds is simple and easy to work with. And on the other it often seems more or less legitimate to set the lack of discrimination that comes with this identification aside, while exploiting the discriminations that the identification does make.

This is also the policy adopted in the model theory of MSDRT, in spite of the fact that propositional attitudes and propositional attitude attributions are its central focus and its raison d'être. After what we have just said about the intensional approach, this may look like an instance of the bizarre incoherence that we just lambasted. But when semantic representation plays as large a part as it does in MSDRT, that offers an escape from the catastrophic effects that lack of content discrimination produces in which meaning can only be accounted for in terms of entities found in models.

For those who are prepared to identify propositions and other non-extensional notions as functions defined on sets of possible worlds – and we have just declared that we have joined their numbers – it is natural to identify the multitudes of possible worlds that the intensional approach presupposes with the worlds needed to account for the difference between a closed past and an open future. That is, all worlds will be histories and some of them will coincide with each other up to some time after which they part company. Or, in terms of models: our models will now be 'doubly indexed' families of extensional models; the indices are now pairs $<w,t>$ consisting of a possible world w and a time t from the history of that world. (Thus for each world w of the set W_M of possible worlds of such a model M the set $\{M_{w,t}$: t is a time from the history of $w\}$ is a history model as defined earlier in this section.) Such doubly indexed families of extensional models will be called *intensional history models*, or simply 'IH models'.

But what is it for t to be 'a time form the history of w'? It seems intuitively plausible that different possible worlds have different time structures. Time, it can be plausibly held, is a product of what happens in a worlds – the product of the events that constitute its history and give it its temporal structure through their relations of precedence and overlap, much in the spirit of the Russell-Wiener construction that was mentioned in Section 1.1. When worlds differ drastically enough in the events they are made up from, then we may expect the instant structures that can be constructed from their event structures to be different too, in their metric or even in their topology.

Allowing for IH models in which each world has its own time structure adds an extra complexities when we want to consider what is the case in different worlds at the same time, for instance when an agent speculates about different possible continuations of the world as it is right now. (What will be the case a year from now when she does A and what will be the case then in case she decides to do B?). For the purposes of this introduction these complications are better set aside. So we will adopt what might be called a 'Newtonian' perspective, according to which time is the same no matter which possible world one is in. That also solves the synchronization problem, by fiat so to speak: Our deliberations now occur at some particular time t of universal time. There is a time t' of universal time that is exactly one year from t and that is a year from now in each of the possible ways in which our worlds can continue after now.[19]

We do assume however that the universal time structures of IH models are always linear orderings. That assumption has consequences for how an IH model can account for open futures. We already hinted at how this can be done: each world is a history with different things happening at the different times of the universal time structure of the IH model. Thus each such world runs from the beginning to the end of time. But certain worlds coincide up to some time from where they do no longer. Formally we capture this relation of coincidence with the help of a ternary relation $\approx$ between two worlds and a time; '$w \approx_t w'$' means that w and w' are indistinguishable at least up to and including t.

This gives us all we need to define the double index structures of IH models. Such a structure is a quadruple $<$W,T,$\prec$,$\approx$$>$, where:

(i) W is a non-empty set (of 'possible histories');

(ii) T is a non-empty set (of 'instants of time');

(iii) $\prec$ is a strict linear ordering of T (the 'earlier-later' relation between the elements of T);

(iv) $\approx$ is a 3-place relation, between a world, a world and a temporal instant. We write: '$w \approx_t w'$', and read this as: 'w and w' are the same world at least up to and including t'.

The formal properties assumed for $\prec$ are that it is a strict linear ordering of T: transitive, asymmetric and total (for all $t, t' \in$ T, $t \prec t'$ or $t = t'$ or $t' \prec t$).

The formal properties of $\approx$ are these:

(i) For each t, $\lambda w.\lambda w'.w \approx_t w'$ is an equivalence relation between worlds;

(ii) If $t \prec t'$ and $w \approx_{t'} w'$, then $w \approx_t w'$.

[19] We could go even more Newtonian by assuming that the structure of time is that of the real number line. But nothing in what will be said in this introduction hangs on this, and we leave it to the reader to think of the universal time structures of our models in these terms.

We are now also ready at last to give a formal definition of intensional history models.

An *intensional history model for* a DRS language LFL is a pair $<<\mathrm{W},\mathrm{T},\prec,\approx>, M>$, where

(i) $<\mathrm{W},\mathrm{T},\prec,\approx>$ is a double index structure; and
(ii) M is a function from pairs $<w,t>$ (with $w \in W$ and $t \in T$) to extensional models $M_{<w,t>}$ for LFL.

Evidently, we have for each world $w \in W$ that the family $\{M_{w,t}\colon t \in T\}$ is a history model, the *history model determined by* w.

Two principles we discussed in connection with history models are (i) Continuity and (ii) the Plenitude of States. Both need to be reconsidered now in relation to IH models. It is intuitively clear that it ought to be possible for one and the same individual **d** to exist in more than one world at the same time even if those worlds do not coincide at that time. Formally:

(1.43) The following is possible: M is an IH model, t an instant from the time structure of M and w and w' worlds from M such that not $w \approx_t w'$ and there is an entity **d** that belongs both to $U_{M_{<w,t>}}$ and to $U_{M_{<w',t>}}$.

Particularly strong examples are those where $w \approx_{t_0} w'$ for some time t_0 before t such that $w \approx_{t_0} w'$ and where **d** belongs to the Universe that the history models M_w and to $M_{w'}$ were still sharing at t_0. Think for instance of what we do when we debate whether to perform action A or action B. When we do A then the world will go on in one direction, when we do B, then it will go in another. Part of what we consider will then in most cases be what the world will be *for us* in each of these cases; we assume unhesitatingly that we will still be there in each of those alternative worlds. And not just us, but countless other things and people that make up the worlds that surround us as we think. That isn't a proof, of course, that things will exist in more than one continuation of the actual world as it is right now, but denying this intuition would seem utterly perverse.[20]

[20](1.43) does not say anything about cases where it is not the case that $w \approx_t w'$ and **d** belongs both to $U_{M_{<w,t>}}$ and to $U_{M_{<w',t>}}$ but there is no earlier time t_0 such that $w \approx_{t_0} w'$ and **d** belongs to the shared Universe of M_w and to $M_{w'}$ at t_0. Plausible cases might be those where it was already determined at t_0 that **d** would come about in the future even though it did not yet exists at t_0 itself. But are there such cases? We leave the question to be resolved by others.

If we acknowledge possibilities like the one described in (1.43), then we should also reconsider the Continuity Principle we stated in (1.33) for history models. Suppose that M and **d** are as described in (1.43). Then continuity should guarantee that **d** also exists at all time between t_0 and t in M_w and likewise that it exists at all time between t_0 and t in $M_{w'}$. But as a matter of fact our statement of Continuity in (1.33) makes sure of this so long as we now regard it as applying to all history models M_w, where w is one of the worlds of an IH model M.

Cross-world identities, where an entity belongs to the Universes that the history models M_w and to $M_{w'}$ were still sharing at t_0 and then continues to exist at times after t_0 when w and w' no longer coincide, will be of special importance when these entities are events or states. This will be so in particular for mental states, for instance about some of your beliefs in the different worlds that would be the result of the different actions you consider: you take it for granted that some of your beliefs won't be affected by what decision you will make – the state that consists in your having such a belief will persist whether you decide this way or that. Here Continuity for IH models will apply as well. The same can be argued for other kinds of states and for events. We leave it to others to argue the point if they see a need for arguing it.

The Plenitude of States principle (see (1.38)) also needs reconsideration in the context of IH models. And this time more is needed than just transferring our old definition to the history models that make up IH models. We just noted that a state can hold not only at different times but also at the same time in different worlds that are not $\approx$-related at that time. By virtue of Continuity such a state **s** will for each world w where it exists at some time occupy a continuous stretch of time (the duration of **s** in w. In the light of this we can talk of the *world-time extent* of **s** in an IH model M, EXT(**s**,M). This is the set of all pairs $<w,t>$ from the index structure of M such that **s** belongs to $U_{M_{<w,t>}}$. Further, let a *continuous region* R of an IH model M be a subset of the set $W \times T$ of its world-time structure with the property that if $<w,t>$ and $<w,t'>$ belong to R and $t \prec t'' \prec t'$, then $<w,t''>$ belongs to R. We can then state the Plenitude of States principle for IH models as follows.

(1.44) Let M be an IH model and **s** a state that belongs to some of its extensional models $M_{<w,t>}$ and that is described by the condition Φ. Let R be a continuous region of M such that $R \subseteq$ EXT(**s**,M). Then there is a state **s'** that can be described by Φ and for which it is the case that EXT(**s'**,M) = R.

This brings us to the end of our discussion of the model theory for DRS languages. The models we will be using in the remainder of this introduction will always be IH models, with all the properties that have been introduced in this last section. It is important to keep this in mind, as we will take those properties for granted in what is coming, and not refer to them much explicitly. But they are there in the background no less, and they are very often important.

1.4 Summary of the Chapter

This chapter could be described as a preliminary to an introduction. It has been written with an audience in mind that is familiar with the outlines of DRT and possibly even with more than that, and in the first part we have proceeded rather rapidly and by means of illustrating examples. Our main concern there was to bring to the fore the original motivations for DRT, emphasizing both the impetus from Tense and Aspect and that from nominal anaphora. Both kinds of phenomena suggest that the form of semantic representations must play a part in accounts of those phenomena, and that on their own the semantic values of sentences and discourses of which these are the semantic representations – the propositions expressed by those sentences and discourses – aren't up to this task.

In the second part of the chapter, starting with Section 1.3 and devoted to the model theory for DRT, we have proceeded in much more formal detail. Our reason for this is that the models to which the considerations of this part lead eventuality, the intensional history models introduced in Section 1.3.3, will be the models also used in our presentation of MSDRT. But IH models have been used before in DRT in contexts other than MSDRT. But in those earlier applications the motivations for and formal definitions of such models have never been spelled out with the degree of detail that they have here, and so it will be useful to have got all that has been said about IH models in this chapter out of the way. Still, while it is right for these details to come here in this preliminary chapter rather landing on the same plate as all that we will have to deal with as part of MSDRT's model theory, it remains true that what we are saying about these models is about DRT in general, and not just as foundation and subpart of MSDRT. Seen from this angle the discussion of the model theory of DRT in this MSDRT introduction is somewhat oddly placed. But it is still better to have it here than nowhere at all.

Chapter 2

Introduction to MSDRT I: Motivations, Representation Formats and Constructions of Logical Forms

2.1 The Motivations for MSDRT

• The original motivation for MSDRT was to provide a more flexible and more fine-grained framework for the semantics of propositional attitudes.

At the time when the first steps towards MSDRT were made (the late nineteen eighties), the widespread way of dealing with propositional attitudes in formal semantics and in the analytic philosophy of language, logic and mind more widely, was limited in two ways.

On the one hand there was a very small repertoire of attitude reports (attitude attributions expressed in natural language). For the most part it consisted of sentences like those in (2.1.a,b), with (2.1.c) reaching the limits of complexity (and going beyond what many accounts could cope with).

(2.1) a. John believes that ϕ,

where ϕ is typically some fairly simple complement clause, also in the present tense, like the matrix verb *believes*

b. John desires that ϕ.

with similar constraints on ϕ as (a).

c. John expects to catch a fish and hopes that Mary will cook it.

These limits are narrow and unnatural from the perspective of actual use.

When we make attitude attributions to others, we often do this in order to make sense of something they have done, or of some noteworthy or puzzling feature of their behavior. That usually involves attributing attitudes to them of different 'attitudinal Modes', such as belief, desire, expectation, doubt, wondering whether and so on. (And there are many more. We will make further distinctions when we need them.)

These different attitudes are then often combined in reasoning, for instance in bits of reasoning that lead from combinations of beliefs and desires to intentions. Even when the reasoning is not spelled out in the report itself, it is often intimated, and the way in which the belief, desire and/or intention contents are specified in the report are usually worded in ways that indicate how the attributee of the report might have reasoned.

• This brings us to the second limitation of the state of the art in the second half of the eighties: the use of intensional semantics to identify the contents of propositional attitudes. This is problematic in all sorts of ways. Perhaps the most telling one is the so-called 'Logical Omniscience' problem:

Suppose that an agent A has two beliefs Bel_1 and Bel_2, and that the contents of these two beliefs are sets of possible worlds ('propositions' according to the identification made in intensional semantics) which are mutually exclusive (in that they have an empty intersection). Since entailment between such propositions is a matter of set-theoretic inclusion – p entails q iff $p \subseteq q$ – the conjunction of Bel_1 and Bel_2 entails in any proposition q whatever, since the empty set of worlds is included in any set of worlds. So A seems committed to the belief in each and every proposition, a form of logical incoherence

But coming to hold two incompatible beliefs is something that can happen in all sorts of ways, without A being aware of the incompatibility and through no logical error of hers. The most widely discussed cases in philosophy are those where Bel_1 and Bel_2 are de facto about the same object **d** while attributing contradictory properties to it. That renders the contents of the two beliefs incompatible, but A may be unaware of this because **d** is represented differently in Bel_1 and Bel_2 and A may be unable to realize that these specifications are specifications of the same individual.

We do not want to lump such perfectly rational agents together with agents who do not refrain from entertaining incompatible beliefs in the face of plain evidence that their beliefs could not possibly be jointly satisfied (as in certain cases of plainly wishful thinking). A minimal requirement for this is that belief contents aren't identified with sets of possible worlds.

• MSDRT started as an attempt to overcome both these problems – the logical omniscience problem and the limited repertoire of attitude reports that can be handled – in the following way.

1. First, MSDRT defines a framework for the description of mental states in a form that is attuned to its application in the semantics of attitude reports. These mental state descriptions are sets consisting of descriptions of constituents of two kinds:

(a) *propositional attitudes*

(b) *entity representations*.

The descriptions of propositional attitudes are pairs $<MOD,K>$, where MOD is a 'Mode Indicator', which specifies the kind of the attitude, i.e. whether it is a belief, a desire and so on, and K is a DRS.

Entity Representations (MSDRT's descriptions of entity representations) have the following form:

$$\Big\langle [ENT, \alpha]\ ,\ K\ ,\ \mathcal{K}\ \Big\rangle$$

The first of the three components of an ER, [ENT, α], is a kind of Mode Indicator. It indicates that this is an Entity Representation and not some kind of propositional representation. α is a discourse referent, the *distinguished discourse referent of* the ER.

The second component of K is a DRS, which provides descriptive information about the entity that is represented by the ER, and the third component $\mathcal{K}$ is a set of 'anchors', which link the ER to the entity represented by it. (More details later)

We refer to the mental state descriptions of MSDRT as *MSDs*.

• Specifying the possible forms of MSDs constitutes the first stage of the definition of MSDRT. The second stage is concerned with how MSDs can be integrated into DRSs. (This integration makes MSDRT into a proper extension of DRT.)

2. The second stage takes the form of adding to the predicate *Att* to DRT.

Att is a special predicate in more than one way. It has four argument positions. The schematic form of a DRS Condition of which *Att* is the main predicate looks like this:

(2.2) s: $Att(\alpha$,MDS,LINKS)

Here α is a dref representing the agent, to whom the Condition in (2.2) attributes a mental state that satisfies the description given by the MSD that fills the following argument position of *Att*.

The first slot is for a state dref s. s represents the state that consists in α being in a mental state of the kind described by MSD.

The final slot is for links between ERs from the MSD in third position and elements outside the given *Att* predication. What these links are for cannot be explained properly at this point.

Att-predications are DRS Conditions. They can be part of the Condition sets of DRSs, in combination with other Conditions from the underlying DRS language to which *Att* is added.

The meaning of 'underlying DRS language' is this. Like DRT, MSDRT is a framework for defining many DRS languages, not just one. Different such languages result from choosing different 'underlying DRS languages' and then extending those to corresponding MSDRT languages by adding the predicate *Att* to them.

Note that when *Att* is added to an underlying DRS language L, the resulting MSDRT language L + *Att* will properly include L by a wide margin. This has to do with the fact that the content-specifying DRSs of Propositional Attitude constituents of MSDs may contain *Att*-Conditions in their turn. This makes it possible to express attitude attributions in which an agent A attributes to an attributee B attributions that B makes to some other attributee C. In particular C may be A, in which case we get the back and forth attributions between interacting agents, something that is important in the context of joint planning and action. We will see how this works in detail later on.)

The occurrence of *Att* Conditions in content-specifying DRSs creates the possibility of recursion that gives MSDRT languages much of their expressive power.

• One use that can be made of the MSDs of MSDRT is as descriptions of the mental states of speakers and hearers – more generally: of language producers (in speech or writing) and the recipients/interpreters of what they write or say.

This makes it possible to analyze linguistic meaning in terms of the effects that spoken and written utterances produce in the mental states of their interpreters.

Such a communication-theoretic approach to meaning in language is an important approach to natural language semantics. MSDRT offers one framework in which this approach can be pursued with formal rigor.

• It is time to look at some examples.

Our first example doesn't have anything directly to do with language. It is about changes in the mental state of an agent A who sees (or thinks she sees) a coin lying in the middle of the road. She thinks it is a gold coin, wants to get it and intends to go to the middle of the road to pick it up.

$$
(2.3)\quad \left\{ \begin{array}{l}
\left\langle BEL, \begin{array}{|c|}\hline x \quad s_1 \quad s_2 \\ \hline n \subseteq s_1 \quad n \subseteq s_2 \\ s_1\text{: gold-coin'(x)} \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array} \right\rangle \\[3ex]
\left\langle DES, \begin{array}{|c|}\hline s_3 \\ \hline n \subseteq s_3 \\ s_3\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\[3ex]
\left\langle INT, \begin{array}{|c|}\hline t \quad e \\ \hline n \prec t \quad e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle
\end{array} \right\}
$$

(2.3) presents the additions to A's mental state that result from her perception and the thoughts triggered by it.

Note in particular the recurrence in the DES and the INT constituents of (2.3) of the dref x that is introduced in the BEL constituent. This renders the represented desire and intention *referentially dependent on* the belief constituent. This kind of mind-internal referential dependency is a common feature of mental states. And it is one of the features of MSDs that complicate the model theory for MSDRT.

The drefs i and (as it is used here) n are special drefs. They only occur in DRSs that describe mental states of agents. i represents the agent from her own perspective (the agent's 'self') and n represents the agent's 'psychological present' – what she experiences as 'now' at the time at which the MSD in (2.3) describes her mental state.

An alternative way of describing these changes in A's mind is given in (2.4). In this MSD it is assumed that the visual experience results in an entity representation in the agent's mind for the thing she perceives (or takes herself to be perceiving).

(2.4)

$$
\left\{
\begin{array}{l}
\left\langle [ENT,x], \begin{array}{|c|}\hline x \quad s_2 \\ \hline n \subseteq s_2 \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array}, \left\{ \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: see}(i,x) \\ \hline \end{array} \right\} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline s_3 \\ \hline s_3\text{:}\mathit{gold\ coin}(x) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle DES, \begin{array}{|c|}\hline s_4 \\ \hline n \subseteq s_4 \\ s_4\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle INT, \begin{array}{|c|}\hline t \ e \\ \hline n < t \ e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle
\end{array}
\right\}
$$

This time the dref x is introduced as part of the ER. So long as this ER succeeds in representing an object – the one that is perceived by the agent when her perception is veridical: there is an actual object there, which is the cause of her perceptual experience – then this takes care of the referential dependency problem mentioned in connection with (2.3). If there is an object in the middle of the road that the ER properly refers to, then the recurrence of x in the contents of the belief, desire and intention constituents of the MSD means that these Propositional Attitudes represent *singular propositions about* this object. And this renders the contents of the three attitude components of (2.4) referentially independent from each other.[1]

(This is not so when the perception is not veridical, but some kind of optical illusion and there is no object that is the perceptual cause of the agent's experience. In such cases the different constituents of (2.4) will have the same significance for the agent as they do when the ER does properly refer.)

But note well: the referential dependency problem overtly displayed in (2.3) is not fully removed by the introduction of the ER in (2.4). If A's perception

[1] a proposition p is *singular about* an object d iff there is a property P, such that any world w, p is true in w iff d has P in w.

is an optical illusion and she is not aware of this, she will use the constituents of (2.4) in her reasoning about the (non-existing) coin in the same way she would have if her perception had been veridical. And her reasoning would have been no less rational in this case as it would have been, had she seen a real coin. To account for this in terms of MSDRT, the same semantics is required that is also needed for MSDs with referential dependencies like those in (2.3).

In short: Adding ERs to the repertoire of MSDRT doesn't do away with the referential dependency problem.

We end this first encounter with the notations used in MSDRT with an example involving *Att*.

(2.5)
$$
\begin{array}{|c|}
\hline
s \quad t \quad a \quad c \\
\hline
t \prec n \qquad t \subseteq s \qquad \text{Agnes'}(a) \qquad \text{coin'}(c) \\
s\text{: } Att \left(a, \left\{ \begin{array}{l}
\left\langle [ENT, x], \begin{array}{|c|} \hline x \quad s_2 \\ \hline n \subseteq s_2 \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array}, \left\{ \begin{array}{|c|} \hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: see}(i,x) \\ \hline \end{array} \right\} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|} \hline s_3 \\ \hline s_3\text{: } \textit{gold coin}(x) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle DES, \begin{array}{|c|} \hline s_4 \\ \hline n \subseteq s_4 \\ s_4\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle INT, \begin{array}{|c|} \hline t\ e \\ \hline n < t\ e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \{< x, c >\} \right) \\
\hline
\end{array}
$$

The DRS in (2.5) represents an attribution to an agent named 'Agnes' (and represented by the dref a) of the property of having been at some time t in the past of the attribution time n in a mental state of the kind described by the MSD in the third argument position of *Att*.

One of the claims made by (2.5) is that the ER from the MSD in the *Att*-predication refers to some object represented in the DRS by c. This is the function of the link <x,c> in the fourth argument position of *Att*. (In this case the fourth argument is a singleton set. In general there can be more links for ERs in the MSD.)

2.2 Constructing DRSs for attitude reports.

Here is a list of some examples of attitude reports. We will look at the construction of DRSs as Logical Forms for some of these sentences..

(2.6) a. John believes that it is raining.

b. John believes/d that it is/was raining.

c. John believes that it is raining and (he) hopes that it will stop raining.

d. John is convinced that there is a mistake in his proof and he is dead set on finding it.

e. John believes that Mary is in Paris.

f. John believes that Mary is in Paris. He would like her to be in Berlin.

g. I was long convinced that Goldbach's Conjecture is true. But now I am no longer quite so sure that it is.

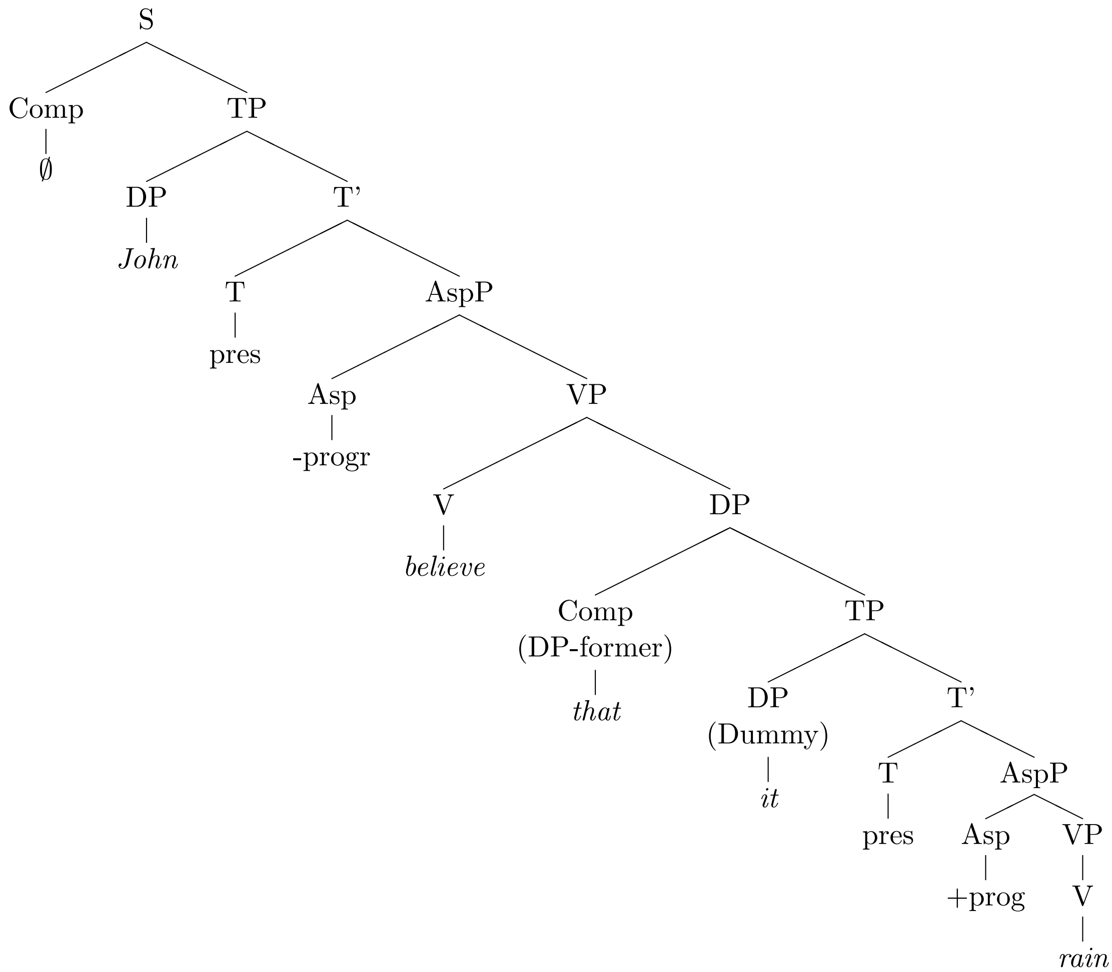
Construction for (2.6.a)

We start with the DRS construction for (2.6.a), repeated below:

(2.6.a) John believes that it is raining.

As usual we need a syntactic structure for the sentence that can serve as input to the DRS construction. For (2.6.a) we adopt the syntactic structure in (2.7).

(2.7)



As we saw in our earlier DRS constructions, our lexicalist approach requires lexical entries for all the content words in the sentences and discourses for which DRSs are to be constructed. New for us at this point are the lexical entries for attitudinal verbs, such as *believe* in (2.6.a).

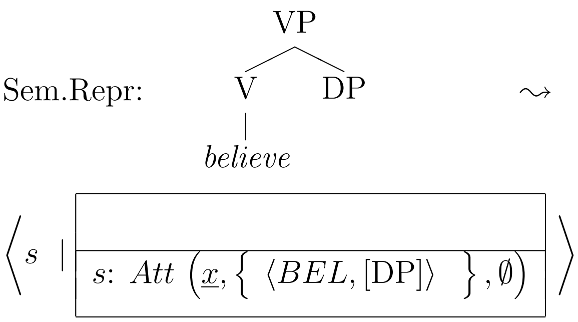
The entry for *believe* is given in (2.8). The entry is unusual, in a way that is specific to attitude predicates. The reason has to do with the *Att*-related structure that attitude verbs must introduce. The convention we will use is to define the semantics of *believe* as an operator which transforms the part of the syntactic tree that is governed by the given occurrence of *believe* – of which the complement/direct object DP is part – into a semantic structure that contains the complement node in square brackets. The meaning of this is that the syntactic subtree headed by the DP node in question is to be substituted as is for the expression '[DP]' in the output structure of the

entry, to be interpreted in the context indicated by its place in the resulting structure. (What that comes to will become clear as we go along.)
(2.8) gives the lexical entry for the verb *believe*.

(2.8) (Lexical entry for *believe*)

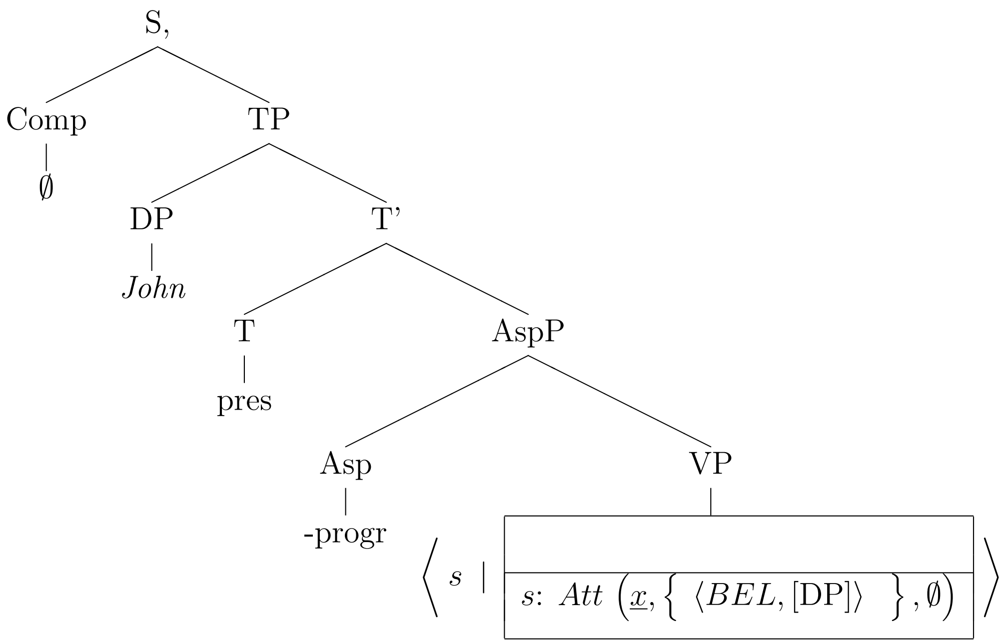
	believe(V_{Att})	nom
	s	$\underline{x}$
Sel. Restr:	state description	agent

Sem.Repr:



The result of applying lexical insertion for *believe* to (2.7) leads to (2.9).

(2.9)



where 'DP' in the expression '[DP]' is short for the subtree of (2.7) that is headed by the second DP node.

The remainder of the construction from the VP up is like what we have seen in our earlier DRS construction. This leads to the representation in (2.10).

(2.10)

$s \quad t \quad j$
$t = n \quad t \subseteq s \quad$ John'(j) s: $Att\Big(j, \Big\{\ \langle BEL, [\mathrm{DP}]\rangle\ \Big\}, \emptyset\Big)$

In general the interpretation of the complement of the attitudinal predicate that occupies the position of the content specification of the corresponding attitudinal Mode Indicator – in (2.10), this is the constituent represented as '[DP]' in the content specification slot for BEL – will depend on the attitudinal context in which it is placed. For the present example there is no such dependence, but that is the exception rather than the rule. In fact, the content DRS of the MSD in (2.10) can be constructed for the most part using construction rules for the underlying DRS language; no rules that are specific to MSDRT are needed in this case).

The first step in the construction of the content DRS of the BEL constituent in (2.10) that we have not yet seen is the transition from the lower T' node to the TP node.

Normally this step inserts the referential argument of the subject DP into the corresponding argument slot of the T' representation. But in the present instance the subject is the dummy subject *it*, and there is no argument slot in the T' representation. The reason for this is that we treat this and other weather verbs as event verbs without additional argument slots, i.e. as 1-place predicates of events. The lexical entry for *rain* is given in (2.11).

(2.11) (lexical entry for the verb *rain*)

	rain (V) e
Sel. Restr:	event
Sem.Repr:	$<e \mid \begin{array}{|c|}\hline \\ \hline e\text{: rain'} \\ \hline\end{array} >$

No argument insertion is possible or needed when the subject is a dummy subject. So the construction rule in this case, triggered by the feature ‘Dummy’ in (2.7), is to pass on the T’ representation unchanged to the TP node.

The next step is specific to MSDRT. The syntactic assumption here is that the particle *that* and the subtree headed by TP can combine to yield a DP ‘that it is raining’. This is motivated by the observation that the verb *believe*, as it occurs in belief-attributing sentences such as (2.6.a,), is a transitive verb, whose direct object argument is restricted to ‘propositional’ entities.[2]. On this view the *that*-complements of attitudinal predicates are definite DPs, which refer to the proposition, or proposition-like object[3] that is determined by the clause that *that* combines with.

The construction step which leads from the TP representation to that of the *that* DP may involve transfer of all or some of the drefs in the TP representation store to the Universe of the DRS to its right. In the present case this applies to the drefs t and s' in the store of the TP representation.

(2.12.a) gives the TP representation of the complement in (2.7) and (2.12.b) the representation of the complement DP. (2.12.c) is the representation of the entire sentence (2.6.a).

[2]In this respect uses of *believe* in which it is combined with a ‘*that*-complement’ are like occurrences that combine with pronouns (’He believes it/that too’), free relatives or certain quantifier phrases (‘He believes whatever she tells him’, ‘I believe some of the things you are saying but not all of them’)

[3]We will eventually adopt *context change potentials* (*ccp*s) as the denotations of the *that*-complements of attitudinal predicates. Context change potentials are prominent in the work of Heim (Heim 1983) and Groenendijk, Stokhof and Veltman (Groenendijk, Stokhof & Veltman 1996*b*). For details see Ch. 4. Section 2.

(2.12)a.

$$<t', s'_{ref} \mid \begin{array}{|c|}\hline \\ \hline t' = n \quad t' \subseteq s' \\ \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \\ \hline \end{array} >$$

b.

$$\begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \\ \hline \end{array}$$

c.

$$\begin{array}{|c|}\hline s \quad t \quad j \\ \hline \\ t = n \quad t \subseteq s \quad \text{John'}(j) \\ \\ s\text{: } Att\left(j, \left\{ \left\langle BEL, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\ \\ \hline \end{array}$$

2.2.1 The last three sentences represented in 2.6.b. Issues of tense and psychological presents.

(2.6.b), 'John believes/d that it is/was raining', is a compressed presentation of four different sentences, one of which is (2.6.b). The other three sentences are listed separately below, as (2.6.b.2,3,4):

(2.6.b.2) John believes that it was raining.

(2.6.b.3) John believed that it was raining.

(2.6.b.4) John believed that it is raining.

The representation construction for (2.6.b.2) is very much like that for (2.6.b). The only difference is that the belief complement DRS now has the form in (2.13).

$$(2.13)\quad \begin{array}{|c|}\hline t' \;\; s' \\ \hline t' \prec n \quad t' \subseteq s' \\ \\ s'\text{: PROG}(\wedge e.\begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \\ \hline \end{array}$$

The only difference between (2.12.b) and (2.13) is that the latter has '$t' \prec n$' where the former has $t' = n$. This difference is due to the tense features at the T nodes of the syntactic structures for the complements of *believe*. In (2.7) this feature is 'pres', in the syntactic structure of (2.6.b.2) it is 'past'. The feature 'past' is responsible for the temporal location Condition '$t' \prec n$'.

We must be careful to properly understand what the Condition '$t' \prec n$' in (2.13) and the corresponding '$t' = n$' in (2.12.b) and (2.12.c) contribute to the representations of the respective belief sentences. The occurrences of the dref n in these Conditions stand for John's 'psychological present' as part of his mental state at the time of the belief that the sentences attribute to him. In this regard these occurrences differ from that of n in the Condition '$t = n$' in (2.12.c). This latter occurrence of n represents the utterance time of the attributing sentence (2.6.a) and in that it is like all occurrences of n we have encountered in Ch. 1. The occurrences of n in (2.12.b) and (2.13) of n are different.

But that is not so in general. It is not so, for instance, for our next sentence (2.6.b.3), where the report is in the past tense and the location time t of the state s that is in the first argument position of Att is in the past of the external now. In this case the psychological now represented by occurrences of n inside the content representation of the attributed attitude coincides with the location time t of s and thus does not coincide with the external now. Occasionally we will introduce subscripts to n to distinguish between those occurrences of n that represent the 'external' present of the utterance time and those that represent the 'internal' present of the attributee at the time of the attitude or attitudes attributed to her.

For an example of this heuristic practice consider (2.14), as an alternative version of (2.12.c).

$$
(2.14)\quad
\begin{array}{|c|}\hline s\quad t\quad j \\ \hline
t = n_e \quad t \subseteq s \quad \text{John'}(j) \\
s\colon Att\left(j, \left\{ \left\langle BEL, \begin{array}{|l|}\hline \quad t'\ \ s' \\ \hline t' \prec t \qquad t' \subseteq s' \\ (t' \prec n_p) \\ s'\colon \text{PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\colon \text{rain'} \\ \hline\end{array}) \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\ \hline
\end{array}
$$

Note that the location time t' of the state s' of raining has been located somewhere in the past of the time t of the attributed mental state. Since in this case n_p coincides with t, the Condition '$t' \prec n_p$' follows as a consequence. In (2.14) this is indicated by the addition of the Condition in parentheses.

The main point to be retained from this discussion is that in representations of attitude attributions the occurrences of n inside the content representations of attributed attitudes nearly always coincide with the location times of the mental states to which the attributed attitudes are said to belong by the attitude attribution; in the case of (2.14) this is the time t. This point will prove to be of central importance in many examples we will have to deal with in this course.[4]

[4]There is a caveat here. Sometimes it may be necessary to refer within a DRS K that

Sequence of Tense

In (2.6.b.3) the matrix verb occurs in the past tense. (2.15) presents the representation construction for this sentence before the start of the interpretation of the complement DP but after all else has been done – the stage, in other words, that corresponds to the construction stage (2.10) for (2.6.a).

(2.15)

$s \quad t \quad j$
$t \prec n \quad t \subseteq s \quad \text{John'}(j)$ s: $Att\left(j, \left\{ \ \langle BEL, [\text{DP}] \rangle \ \right\}, \emptyset\right)$

How are we to complete the construction for (2.6.b.3) after it has reached the stage shown in (2.15)? An answer that might easily come to mind would

represents the mental state that a person B is in at some time t back to the psychological present of a person A who makes the attribution to B whose content is represented by K. In such cases a symbol must be used in K from which it is clear that it refers to A's psychological present at the time when she made her attribution to B, and not the psychological present of B at the time he is supposed to be in the mental state represented by K according to A's attribution. Such cases are not uncommon, but they will play no direct role in the examples we will discuss in this Introduction.

We also add a general point about notation. Since the two different occurrences of n mentioned above do such different jobs, it might be thought that different symbols should be used for them. But unfortunately having just two different symbols won't be good enough. Among the DRSs that are made available by MSDRT there are those that represent attributions whose content is that the attributee attributes a certain attitude to some third person; and th content of that attitude may in its turn be that some fourth person has some attitude, and so forth. Such representations will have occurrences of n for each of the persons involved. If we want distinct symbols for the psychological presents of each of them, then we would need at least three just for that, in addition to a symbol for the dref representing the utterance time. But there is no upper bound to the number of nested attributions can be involved in such attributing sentences. So in principle we would need infinitely many different symbols.

But then, which of those symbols should be used when and where in the constriction of the DRSs for such attributing sentences? Clearly, the role played by the occurrences of n in (2.12.c) can be inferred from the 'box' (i.e. main DRS and its various sub-DRSs) where n occurs. (More precisely: that DRS that contains the DRS Condition in which n occurs as one of the members of its Condition Set.)

There is a caveat here, however. Sometimes it may be necessary to refer within a DRS K that represents the mental state that a person B is in at some time t back to the psychological present of a person A who makes the attribution to B whose content is represented by K. In such cases a symbol must be used in K from which it is clear that it refers to A's psychological present at the time of her attribution to B rather than the psychological present of B at the time he is supposed to be in the mental state represented by K according to A's attribution. Such cases are not uncommon, but they will play no direct role in the examples we will discuss in this Introduction.

be that the different pieces of the answer are already spread out before us: the complement clause of (2.6.b.3) is identical with the complement clause of (2.6.a) whose representation is shown in (2.13). To complete the representation for (2.6.b.3) all that needs to be done, one might be inclined to think, is to just replace '[DP]' in (2.15) by (2.13).

But that won't quite do. Note that the location time t' of the raining state s' is in the past of the time represented by n in (2.13). That time is the time of the belief state s, and in the representation that we would obtain for (2.6.b.3) if we substitute (2.13) for '[DP]' in (2.15) s is itself located at the time t, which is in the past of the utterance time represented by the n of (2.15). That is, the time t' of s' is, according to the representation for (2.6.b.3) that we obtain by mere substitution of (2.13) for '[DP]', situated in the past of the belief time t. Such an interpretation of (2.6.b.3) cannot be fully ruled out, but it is not the salient one for pretty much every competent speaker of English. The salient interpretation of (2.6.b.3) is that in which s' is simultaneous with s.

This observation is an old one and it has been a central concern of the theory of tense and aspect for many years. The simultaneity of s' and s for (2.6.b.3), and like sentences in which both the attitude verb and the verb of its complement are in the past tense, is one that English shares with a good many others, but it is not a linguistic universal, not even among those languages that mark tenses morphologically. The languages for which it holds are known as 'sequence of tense' languages. In these languages tenses of complement clauses to verbs that take clausal complements can be 'temporally anaphoric' to the states or events described by those verbs, when the tenses in main clause and complement clause agree. (For instance both are past tenses or both tenses are present.) This is not to say, however, that languages that have morphological tense marking but are not sequence of tense languages according to the standard definition, do not have very similar properties of their own. Examples of such languages are Japanese and Russian. They too can express simultaneity of the state described by the complement clause with the eventuality described by the matrix verb; but they express simultaneity with the eventuality introduced by a past matrix verb eventuality by putting the verb of the complement clause into the present tense. This shows that sequence of tense languages need a special rule for the interpretation of 'past under past' configurations.[5]

[5]For work on how simultaneousness is expressed in sequence of tense languages like English and 'non-sequence-of-tense languages like Japanese see in particular the work by

This isn't the place for a proper discussion of sequence of tense. But we need a construction rule to deal with sequence of tense sentences like (2.6.b.3). The first question is at what point of the DRS construction this rule should apply. The answer to this question seems quite clear: its point of application should be that when the T node of the complement clause is combined with its AspP representation. This is the point when the past tense of the complement clause should be able to trigger the rule, and at a point where it is clear whether the conditions for its application are satisfied: (i) that the clause to which the given T node belongs is the complement clause of a clause-embedding verb, (ii) that the matrix verb is in the past (these are just matters of syntactic structure, which we assume have been recognized and made visible by the syntactic parser that is presupposed by our DRS construction algorithms throughout) and (iii) that the complement clause describes a state (as opposed to an event), something that the representation construction for the sister node to T will have established at this point.

The preconditions of the rule and the transformation of the semantic representation of its sister node are given by the following formulation:

(2.16)(Sequence of Tense Rule for 'Past under Past', first formulation)

Suppose that t' is the location time of the state described by the past complement clause of an attitude attributing verb V in the past tense and that t is the location time of the eventuality described by V. Then set t' equal to t.

Temporal Deixis and Presupposition

But this isn't the form in which we want to state the rule. (2.16) treats the embedded tense as anaphoric to the temporal location of the matrix verb, and thus as anaphoric in some sense. That is consistent with a feature of tenses more generally, but one that we haven't so far touched on here. So far we have treated the past tense as an existential quantifier: *It rained* is true now if and only there was some time t before now such that it rained at t, and so on. That idea goes back to the earliest proposals for the semantics of the tenses (as for instance in the work of Prior (Prior 1967)). But that this isn't right was shown early on in a short but seminal paper of Partee (Partee 1973). The central example of this paper is the sentence (2.17), said by one half of the couple who have just tied up and left the house for a long trip.

Ogihara, e.g. (Ogihara 1992), (Ogihara 1994), (Stowell 2007), **more references?**.

(2.17) I didn't turn off the stove.

It is intuitively obvious in the given situation that what (2.17) expresses is that the speaker forgot to turn off the stove while they were getting ready for the trip. Unless this constraint on the domain of temporal quantification is made part of what the past tense *didn't* expresses in (2.17), neither of the two possible scope relations between the existential quantifier contributed by the past tense and the negation will deliver an interpretation that comes near to this intuitive meaning of the sentence.

One response to Partee's observation is that the existential quantifiers expressed by tenses are subject to Domain restriction, much like this has come to be realized in relation to nominal quantifiers like *every*, *most* or *no*.[6] Furthermore, it is natural to see domain restriction as a form of anaphoric presupposition: the restricted domain has to be recovered from the local or wider context – a presupposition triggered by the quantifying expression. In the case at hand the presupposition triggers are the tenses.

Such a presuppositional treatment of the tenses is proposed in (Kamp et al. 2011). In particular, the general format for the presuppositions triggered by tenses is as in (2.18).

(2.18)
$$\begin{array}{|c|}\hline r?\ \ \rho? \\ \hline \rho(r,t') \\ \hline \end{array}$$

The presupposition representation in (2.18) is doubly anaphoric: identifications must be found both for the 'queried dref' '$r?$' and for the 'queried dref' '$\rho?$'. $\rho?$ must he instantiated by a temporal relation (such as $\subseteq$, $\prec$, $\supset\subset$ or $=$); $r?$ must be instantiated by a time (some interval of the time axis). As in (Kamp et al. 2011), we assume that presuppositions of this form are triggered by all T features and that the meaning of the individual tense features includes constraints on how the queried drefs are to be resolved in the context. In the case of (2.6.b.2) this applies both to the feature 'past' of the main clause and that of the complement clause. But of course it is the 'past' feature of the complement clause that matters to us right now. Note that the resolution rules for the two past tenses must be different. The main clause tense must locate the location time t of the eventuality described by the main clause verb in the past of the utterance time n, via the Condition 't

[6]For an often cited publication on Domain restriction see (Stanley & Szabo 2000).

$\prec n$'. (How the presupposition schema (2.18) is to be resolved in this case is a question that this isn't the place to say more about.) But the past tense of the complement should not trigger the Condition '$t' \prec n$', since that would locate the state described by the content of the attitude as holding before the psychological 'now' n_p of the attributee, which is not what we want here. Instead the presupposition resolution we want here is 'strictly anaphoric': $r?$ is to resolved to the 'anaphoric antecedent' – for (2.6.b.2) this, as we noted, the location time t of the main clause eventuality – and ρ is to be resolved to $=$.

This is what we are told by the anaphoric tense rule in (2.16). So this rule should be stated in the form of a resolution constraint on the presupposition that is triggered by the particular past tense feature of the lower T node in the syntactic structure for (2.6.b.2). I will not rewrite the rule in this spirit here, as that would require us to set up the regime for the presuppositional dimension of the tenses in greater formal detail than seems reasonable here. But let us assume that the T feature that triggers this anaphoric interpretation of the contribution by tense is 'past$_{mvt}$' ('mvt' is short for 'matrix verb time').

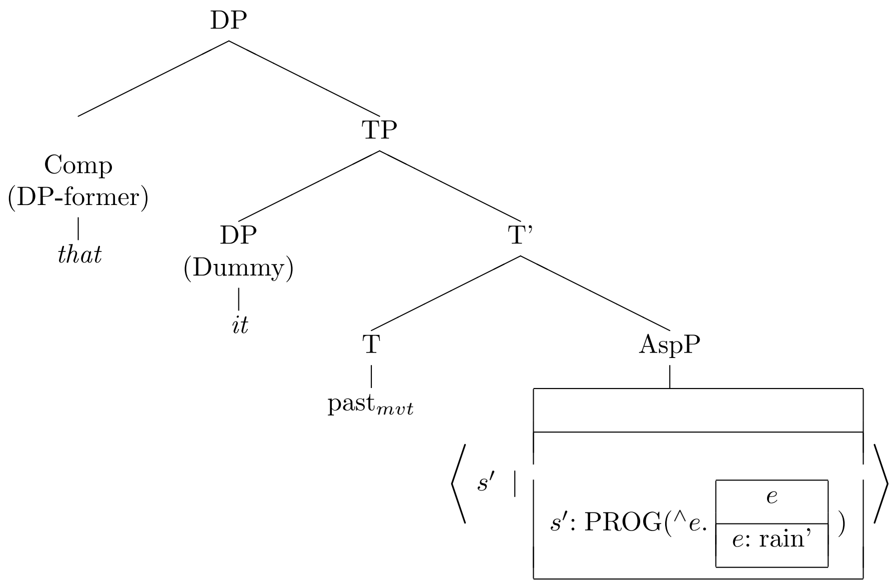
To know how anaphoric past tenses like the one in the complement clause of (2.6.b.2) are to be resolved is one thing. The other thing is which occurrences of the past tense should be resolved in this manner. The possibility of such an interpretation depends on the one hand on the syntactic position of the verb that's bears the tense – the given interpretation is in general possible only for past tenses of the main verbs of complement clauses. This is information that should be recognizable by the parser that provides the syntactic structures from which DRSs are constructed. So we assume that it is the parser that decorates the relevant T nodes with the part tense feature that triggers this interpretation of tense. But there is also a semantic constraint on the interpretation we are considering: the complement clause must be the description of a state (and not of an event). This information cannot be assumed to be available to the parser. The parser has marked the given occurrence of T with the feature that triggers the simultaneity interpretation and the complement representation turns out to be an event description, then the derivation crashes.

Although we will not spell out a fully worked-out presuppositional treatment of the tenses here, we will need the general format for representing presuppositions before they are resolved, so that their resolution can be formally defined as the resolution of those representations. The general architecture for the construction of semantic representations proposed in (Kamp et al. 2011) follows (Van Der Sandt 1992) in splitting the construction into two stages: first a *preliminary* DRS is constructed, in which the presuppositions triggered by presupposition triggering elements of the sentence of utterance for which the representation is being constructed are explicitly represented. The second stage consists in the *resolution* of the presuppositions represented in the preliminary representations, fully on the basis of the available context information or otherwise by means of full or partial accommodation (Beaver 1997), (Beaver 2001). (There are also cases when a presupposition cannot be resolved because the needed accommodation conflicts with commitments that cannot be given up. In such cases the construction aborts and no viable interpretation is forthcoming.) When a presupposition that is represented in the preliminary representation has been resolved, the presupposition representation is eliminated from the preliminary representation. When all presuppositions have been resolved, the representation is of the form we have been assuming here so far: a DRS preceded by a store (or, when the store is empty, just a DRS). When the sentence for which a representation is being constructed has no presupposition triggers, as we have been assuming here up to now, then the two stages collapse into one: there are never any presupposition representations to begin with and so there is no difference between preliminary representation and final representation.

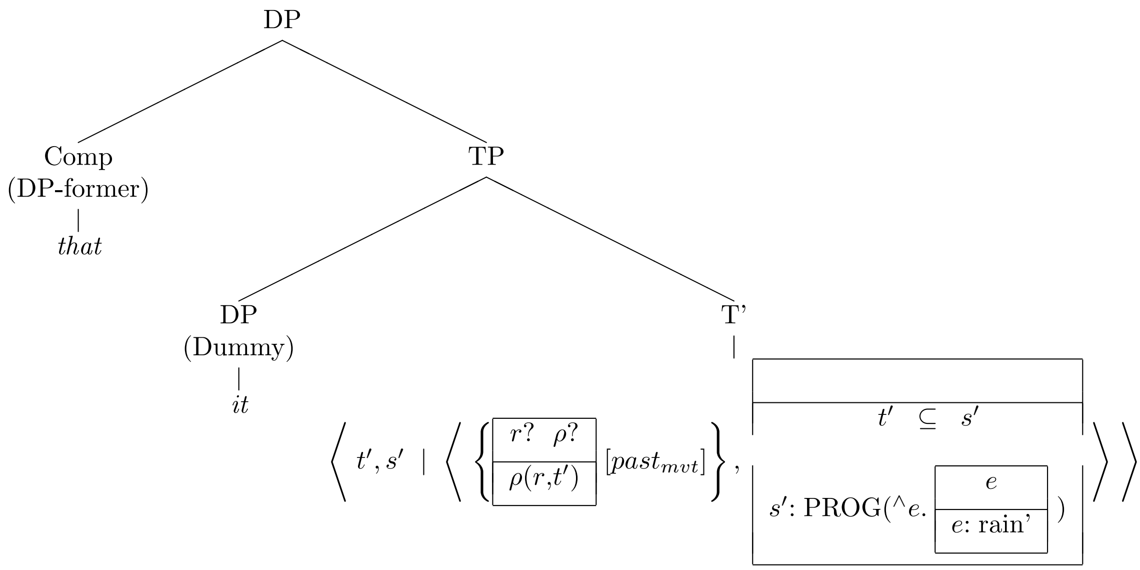
From now on we assume this two-stage architecture, though we will often simplify constructions when presuppositions triggered by the sentence for which we are constructing a representation are irrelevant to the point that the construction is intended to illustrate; such presuppositions we will resolve on the fly, without representing them explicitly.

The two stage approach to presupposition representation and resolution, and the notation used in the presentation of preliminary representations are shown in (2.19). (2.19.a) shows the syntactic structure for the complement clause of (2.6.b.3), with the T feature and with the representation for the AspP node already in place. (2.19.b) shows the result of applying the operation triggered by past_{mvt} to the AspP representation, in the form of adding a version of (2.18) as presupposition representation. (The subscript 't_{mvt}' of 'past' in (2.19.a) stands in for the constraints on presupposition resolution that are described in (2.18).) The set brackets in (2.19.b) indicate that the slot between store and DRS is for a set of presuppositions; in the present case this set is a singleton, but in general it is not. (2.19.c) is the DP representation that results from continuing the construction of (2.19.b).

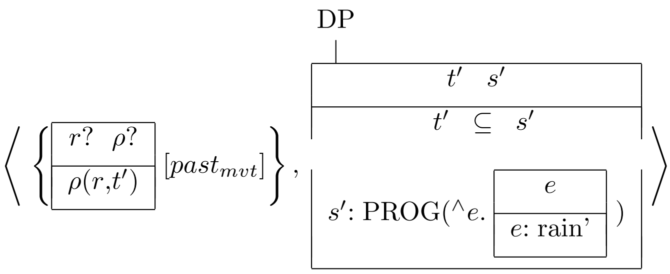
(2.19)a.


DP
Comp
(DP-former)
that
TP
DP
(Dummy)
it
T'
T
past$_{mvt}$
AspP
s'
s': PROG($^\wedge e.$
e
e: rain'
)

b.


DP
Comp
(DP-former)
that
TP
DP
(Dummy)
it
T'
t', s'
$r?\ \ \rho?$
$\rho(r,t')$
[$past_{mvt}$]
$t' \subseteq s'$
s': PROG($^\wedge e.$
e
e: rain'
)

c.


DP
$r?\ \ \rho?$
$\rho(r,t')$
[$past_{mvt}$]
$t'\ \ s'$
$t' \subseteq s'$
s': PROG($^\wedge e.$
e
e: rain'
)

(2.19.c) is part of the preliminary representation for (2.6.b.2) shown in (2.20).

(2.20)

$$
\begin{array}{|c|}
\hline
s \quad t \quad j \\
\hline
t \prec n \quad t \subseteq s \quad \text{John'}(j) \\
s\text{: } Att \left(j, \left\{ \left\langle BEL, \left\langle \left\{ \begin{array}{|c|} \hline r?\ \rho? \\ \hline \rho(r,t') \\ \hline \end{array} [past_{mvt}] \right\}, \begin{array}{|c|} \hline t' \quad s' \\ \hline t' \subseteq s' \\ s'\text{: PROG}(\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \right\rangle \right\}, \emptyset \right) \\
\hline
\end{array}
$$

The next step is the resolution of the presupposition of (2.20). To repeat, resolution of this presupposition must involve (i) resolution of r to the location time t of the state described by the *Att*-predication containing the MSD; (ii) resolution of ρ to =. The representation that results from these resolutions is the one shown in (2.21).

$$
(2.21)\quad
\begin{array}{|c|}
\hline
s \quad t \quad j \\
\hline
t \prec n \quad t \subseteq s \quad \text{John'}(j) \\
s\text{: } Att \left(j, \left\{ \left\langle BEL, \begin{array}{|c|} \hline t' \quad s' \\ \hline t' = t \quad t' \subseteq s' \\ s'\text{: PROG}(\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\
\hline
\end{array}
$$

This gives us the prominent reading of (2.6.b.3) and with it an answer to our first question about the semantics of this sentence. The second question I raised was whether this is its only possible reading. We already noted that other readings are possible as well, and, more strongly, that only another reading is possible when the complement clause describes an event rather than a state. This point is illustrated by (2.22.b). (2.22.c) shows that a non-simultaneous interpretation is possible also when the complement clause describes a state. But with state-describing complements a special context is needed to bring such a non-simultaneity interpretation to the fore. (In (2.22.c) this is the context that the first sentence provides for the interpretation of the second.) In the absence of such a special context the default interpretation (of the complement state being simultaneous with the matrix verb eventuality) seems almost impossible to overrule.

(2.22)a. John said/knew/didn't know that Mary had a new car.
(Simultaneity reading possible and strongly preferred)

b. John said/knew/didn't know that Mary bought a new car.
(Simultaneity reading impossible)

c. That recital was one of Elly's most staggering achievements. Everyone who came to her dressing room afterwards told her she was absolutely brilliant.

The second sentence of (2.22.c) is ambiguous between two readings. On the first, the comments of the people who came to Elly's dressing room after the recital was over were about her as a singer, whose brilliance was demonstrated by her performance, but extending over a much longer period than just the recital; interpreted in this way the second sentence of (2.22.c) is an instance of the simultaneity interpretation of sentences whose matrix verbs have complements that are state describing. On this interpretation the past tense of the second sentence is interpreted by using rule (2.18). But it is also possible to give a different interpretation to the second sentence, according to which it was Elly's performance at the recital of which she was told that it was brilliant. When interpreted this way the state described by the complement clause of *told her* is located in the past of the telling and the interpretation strategy is like that which places the event described by a complement clause in the past of the matrix verb eventuality.

In this respect (2.22.c) is in our opinion genuinely ambiguous. There are also examples of sentences with past tense state describing complement clauses in which the state described by the complement can only be reasonably understood as in the past of the matrix eventuality.[7]

---

[7]The following is an example:

(2.23) When I asked them about it, John thought that they had been there all alone, but Mary thought that there was someone else there as well.

How should we deal with the temporal ambiguity of a sentence like (2.22.c)? Given the road on which we have already embarked by assuming the special T feature past$_{mvt}$ for the simultaneity interpretation, the natural way to continue is to assume a further special T feature for past tense complement clauses that triggers interpretations in which the complement clause is located in the past of the matrix verb eventuality. But this is only one further question about the uses of tenses in attitude attributions. There are any more such questions. They form an important chapter of the syntax-semantics interface of attitude attributing sentences (and in indirect discourse generally). But the answers to those questions can vary substantially from language to language. That makes answering them an even harder and more comprehensive task, which goes beyond the present Introduction, which only consider attitude attributions in English. So we will cut what more could and should be said about this topic short, putting most of the little that we will say on it here in a couple of footnotes, which the reader may read or leave.[8]

Here the time of the state described by 'there was someone else there as well' is that of the time of the episode that the speaker asked John and Mary about and that time is before the time at which John and Mary had their respective thoughts according to (2.23).

[8]Prominent among the combinations of tenses in matrix verb and complement clause is that of simple past for the matrix verb and past perfect for the complement. Except in special contexts this combination convey that the thought attributed to the attributee is about the attributee's past at the time of the thought. That makes a sentence like (2.24.a) truth.conditionally to one of the sentences in (2.22.b)), which we repeat here as (2.24.b). (The interpretation mechanisms involved, for the embedded simple past and the embedded past perfect, which are responsible for this,are significantly different, a point we note without going into details.)

(2.24) a. John said/knew/didn't know that Mary had a new car.
(Simultaneity reading possible and strongly preferred)

b. John knew that Mary had bought a new car.

c. John knew that Mary bought a new car.

Another Important feature of what tenses should be used to describe the contents of attributed thoughts concerns thoughts about the future. Consider the sentence in (2.25), a variation of an example from Abusch (Abusch 1997).

(2.25) John thought last week that in three weeks' time he would take Susan out to dinner and tell her that he loved her.

There are three points that are illustrated by this example:

(i) In order to describe the content of a thought as prospective (i.e. as about the future from the thinker's perspective at the time of the thought) some form of future tense must be used. In (2.25) neither the present tense *takes* nor the past tense *took* would have done in lieu of the past future form *would take*.

2.2.2 A report of two attitudes with different attitudinal modes

This section is about report (2.6.c), repeated here:

(2.6.c) John believes that it is raining and (he) hopes that it will stop raining.

(2.6.c) is a compressed representation of two sentences, one with and one without the pronoun *he*. We will refer to the one with the pronoun as (2.6.c.1) and to the one without the pronoun as (2.6.c.2):

(2.6.c.1) John believes that it is raining and he hopes that it will stop raining.

(2.6.c.2) John believes that it is raining and hopes that it will stop raining.

First (2.6.c.1). The representation for the first conjunct of (2.6.c.1) has already been constructed, as representation for (2.6.a). We repeat the final outcome, given earlier as (2.12.c).

(ii) Although the episode that John was thinking about is located in the future of the utterance time (because of the temporal adverbs *last week* and *in three weeks time*), the past future *would take* is possible here and may even be preferred over the simple future *will take* (for reasons we don't go into).

(iii) Once a future tense has made it clear that the reported thought is about the future, it is then possible to continue using the simple past. This is seen in (2.25), where the simple past tense *loved* is understood as referring to the same time as that of telling Susan. (This use of the simple past is another instance of English sequence of tense, which is operative not only with past-under-past but also with past-under-past future.

The last variant in of the sentences in (2.6), the present-under-past sentence (2.6.b.4) presents another challenge. English present-under-past sentences are known in the tense-and-aspect literature as cases of 'double access'. In sequence of tense languages like English the present tense of the complement of a part tense attitude verb means that the complement describes a state and that this state temporally includes both the utterance time of the attributing sentence and the earlier time of the reported thought.

Both the phenomena illustrated by (2.25) and the double access phenomenon have had extensive attention in the literature. For double access see e.g. (Enç 1986), (Ogihara 1999), (Bary & Altshuler n.d.). For a helpful overview of embedded tenses in English and some other languages see (Ogihara & Sharvit 2015).

(2.12.c)

$$
\begin{array}{|c|}
\hline
t \quad s \quad j \\
\hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
s\text{: } Att\left(j, \left\{ \left\langle BEL, \begin{array}{|l|} \hline \quad t' \;\; s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline\end{array}) \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\
\hline
\end{array}
$$

In most respects the construction of the representation for the second conjunct of (2.6.c.1) proceeds in the same way. But since our discussion of the construction of (2.12.c), one assumption that we made as part of the construction of (2.12.c) has changed. When constructing the representation for (2.6.b.2), we switched to a treatment of tense features (the features determined by T nodes) in which these trigger presuppositions. That assumption was needed there for the past tense of the complement clause, but we now need it also for the treatment of the main clause (i.e. the present tense of the verb form *hope*).

The VP 'hope that it will stop raining' gives rise to a 'hope representation': a pair of the form $<HOPE,K>$ that is a constituent of an MSD that occupies the third argument slot of an *Att* Condition which in its turn belongs to the Condition Set of the DRS that we are constructing. (Responsible for the construction of this structure is a construction principle associated with the verb *to hope* when used, as it is here, with a *that*-complement – a principle that is like the principle associated with the verb *believe* that we have used in the earlier constructions of this section:

The *Att*-Condition that results from applying this rule to the *hope*-clause of (2.6.c.1) has the form shown in (2.27). (Once again we have had to split this representation into two parts, because we couldn't make the single structure fit on the page.). (2.27.a) shows the full representation of the preliminary representation of the *Att*-Condition, in which the pronoun that provides its second argument still has to be resolved and the state dref s'' in first argument position still needs to be bound in some way and thus is still in store.

The DP constituent is shown in full in (2.27.b).[9]

(2.27)

a. $$\left\langle t'', s''_{ref} \mid \left\langle \left\{ \begin{array}{|c|}\hline z? \\ \hline \text{male}(z) \\ \hline \end{array}\, , \quad \begin{array}{|c|}\hline r?\ \rho? \\ \hline \rho(r,t'') \\ \hline \end{array}\,[pres] \right\}, \begin{array}{|c|}\hline \quad \\ \hline t'' = n \quad t'' \subseteq s'' \\ s''\text{: } Att\Big(z, \Big\{ \ \langle HOPE, [DP]\rangle \ \Big\}, \emptyset\Big) \\ \hline \end{array} \right\rangle \right\rangle$$

[9]We will not have anything of substance to say here about the semantics of *hope*, as distinct from *believe*, *desire* or *want*. But here, by way of a first throw in the direction of a proper account: The verb *hope*, when combined with a *that*-complement, expresses a propositional attitude that on the one hand has a doxastic dimension to the effect that one cannot hope that something will be the case (or is or was) and at the same time be 100% convinced of the contrary. At the same time to hope that q seems to imply that the one doesn't take it for granted that q.
It may be a point of debate whether hope excludes belief. If you are certain that Bill will call tonight, does that preclude you from having an attitude that can be described as your hoping that Bill will call tonight? The following example throws some light on this question. Compare (2.26.a) with (2.26.b) and (2.26.c).

(2.26) a. Sue is certain that Bill will call tonight. And she hopes that he will.

b. Sue is certain that Bill will call tonight. And she is happy that he will.

c. Sue is certain that Bill will call tonight. And that is just what she wants.

For us (2.26.a) is less felicitous than (2.26.b) and (2.26.c), This suggests that hope is not happily compatible with certainty, in contrast with the attitudes expressed by the second sentences of .(2.26.b) and (2.26.c). But the judgments are delicate and we fell more canvassing is needed.
hope obviously also has an emotive dimension, which makes it more like a desire or want: *hope* conveys that its grammatical subject would like the proposition expressed by the complement to be true. (*fear* is just the opposite of *hope* in this regard.) But even though *hope* resembles verbs like *want* and *desire* in expressing a positive attitude towards the content of its complement, it doesn't have a volitional dimension. When you ask me to explain why A did something, I can answer your question be saying that she wanted to do this, or that she had been wanting to do this for a long time. But it would be strange for me to reply that she did it because she hoped that she would do it. Related to this is that it is odd to use the verb *to hope* in connection with something over which one has total control. It is coherent to say something like: "I hope I will write this paper", or also "I hope I will manage to write this paper", for these are things that one might fail to do in spite of one's good and serious intentions, or that might fail with even though one has a serious go at them. But in the absence of some quite special context it seems quite odd for some one to say "I hope I will promise to write this paper". If someone did say this to you, your natural reaction would probably be "Well, why don't you then?'.

$$\text{b. DP} = \left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline s''? \\ \hline s'': \text{PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array}\ , \\ \\ \begin{array}{|c|}\hline r'?\ \rho? \\ \hline \rho'(r',t''') \\ \hline \end{array}\ [fut] \end{array} \right\}, \begin{array}{|c|}\hline t''' \quad e \quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'',s^4) \\ (s'' \supset\subset e \supset\subset s^4) \\ s^4\!: \begin{array}{|c|}\hline \ \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s^4 \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle$$

Legenda for (2.27.b):

1. The pronoun *he* is treated as trigger of an anaphoric presupposition. The only constraint it imposes on its anaphoric antecedent is that it must stand for a male individual.

2. The verb *stop* is also treated as a presupposition trigger (a standard assumption). *stop* operates on the representation of its complement – here this is the gerund '(it) raining'. It imposes on its complement the constraint that it must be the description of a state. Its contribution to the semantic representation consists of a presupposition and a non-presuppositional part. The presupposition is that over some interval of time abutting the location time t'' introduced by tense there was a state s'' of the description provided by the complement, and the non-presuppositional contribution is that at t'' this state comes to an end and is followed by a state s^4 where the complement does not hold. (s^4 is a state during which there is no state of the kind described by the complement.)

One way in which this transition from s'' to s^4 can be understood is as the occurrence of a transition event. That is what has been assumed in (2.27.b). the event e is the transition between the two states, with the temporal implication that s'' abuts e and e abuts s^4. In (2.27.b) I have added this implication in parentheses, to indicate that it can be inferred from the Condition 'e: Trans(s'',s^4)', rather than a direct contribution by the DRS construction. Temporal abutment is represented as '$\supset\subset$'.[10]

3. For the sake of uniformity the future tense of the complement clause of (2.6.c.1) has also been treated as presupposition trigger.

[end Legenda]

The representation in (2.27) is a preliminary representation, with a number of presuppositions that need to be resolved. Note that not all these presuppositions are attached in the same position, directly to the left of those DRSs that represent the parts of the input LF that contain their triggers.[11]

We now come to the crucial point about the interpretation of (2.6.c.1). It is one that applies to many attitude reports that attribute more than one attitude to the same agent at the same time:

Often different attitude attributions are presented as parts of the same mental state, belonging to some attributee at some particular time. In such cases, the semantic representation of one attitude attribution is often needed as discourse context for the interpretation of the next part. In the present case we need the content of the belief described in (2.12.c) in order to resolve the *stop*-presupposition of (2.27.b).

But how precisely can we justify the availability of the belief content representation in (2.12.c) for this purpose? At an informal level the justification has already been given: in mental states described by complex MSDs which consist of more than one propositional attitude it is possible for one attitude content to build on that of one or more others. That was the original moti-

[10]In (2.6.c) the dummy subject *it* has been moved to the subject position of the 'raising verb' *stop*. Raising phenomena are not dealt with explicitly here, but we assume that the semantic representation of 'raining' in the second conjunct of (2.6.c) is computed in the same way as that of 'is raining' in its first conjunct. One way to make this explicit would be to reconstruct the dummy *it* as dummy subject of the gerundive 'raining' in the LF from which the semantic representation is constructed. Furthermore, that the resulting representation is like the representation of the progressive rather than the non-progressive form, I take to be an instance of coercion by *stop*, which requires a state description as input. (Note that in 'It rained but then it stopped', 'it stopped' is also short for 'it stopped raining'. In general, though, gerundive forms can be used both to describe states and to describe events.)

[11]For more on this see (Kamp 2021*a*), Ch. 4.

vation for MSDRT to start with; recall diagram (2.3) on page 67.

Part of the answer to this question is given by (2.28). We state it here without further justification, trusting it is plausible enough as stated.

(2.28) When a belief attribution and a hope attribution are constituents of one and the same MSD, then the interpretation of the hope attribution may make use of the content representation of the belief attribution.

If (2.28) is to be of any use for the resolutions of the presuppositions in (2.27.b), however, then we need a further principle that enables us to unite the hope constituent of (2.27.b) and the belief constituent of (2.12) into a single MSD. That this should be possible is also clear intuitively. But a number of construction steps are needed to get to that point. The first step is to merge the representation for the hope conjunct with the representation of the belief conjunct, which according to (2.28) can serve as discourse context for the hope conjunct. (This is standard DRT, except that we merge the two DRSs before having resolved the presuppositions of the hope conjunct representation; but this is not an important difference.) The result is shown in (2.29).

(2.29)

$$\left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline z? \\ \hline \text{male}(z) \\ \hline\end{array}\,, \\ \\ \begin{array}{|c|}\hline r?\ \rho? \\ \hline \rho(r,t'') \\ \hline\end{array}\,[pres] \end{array} \right\}, \begin{array}{|c|}\hline s \quad t \quad j \quad t'' \quad s'' \\ \hline t = n \quad t \subseteq s \quad \text{John'}(j) \\ t'' = n \quad t'' \subseteq s'' \\ s\text{: } Att\left(j, \left\{ \left\langle BEL, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline\end{array}) \\ \hline\end{array} \right\rangle \right\}, \emptyset\right) \\ s''\text{: } Att\left(z, \left\{ \langle HOPE, [DP] \rangle \right\}, \emptyset\right) \\ \hline\end{array} \right\rangle$$

where [DP] is once more the structure (2.27.b).

The next step is to resolve the presuppositions in (2.29). The pronoun presupposition can be resolved to the dref j for John and the temporal presupposition by setting r equal to n and ρ to the identity relation =. These resolutions turn (2.29) into (2.30).

(2.30)

$$
\begin{array}{|c|}
\hline
s \quad t \quad j \quad z \quad r \quad t'' \quad s'' \\
\hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
z = j \quad r = n \quad r = t \quad t'' = n \quad t'' \subseteq s'' \\
s\text{: } Att\left(j, \left\{ \left\langle BEL, \begin{array}{|c|}
\hline
t' \; s' \\
\hline
t' = n \quad t' \subseteq s' \\
s'\text{: PROG}(\hat{}e. \begin{array}{|c|}
\hline
e \\
\hline
e\text{: rain'} \\
\hline
\end{array}) \\
\hline
\end{array} \right\rangle \right\}, \emptyset \right) \\
s'' : Att\Big(z, \Big\{ \ \langle HOPE, [DP] \rangle \ \Big\}, \emptyset\Big) \\
\hline
\end{array}
$$

We are now in a position to merge the MSDs of the two *Att*-predications in (2.30) into a single MSD. This is because, as (2.30) shows, these *Att*-predications describe the mental state of the same agent (viz. John) at the same time. What we are appealing to is the intuitively unimpeachable principle that two descriptions of the same mental state can be merged into a single description. The result of making a single *Att*-predication of the two in (2.30), with a single MSD that is the union of the MSDs of the two *Att*-predications (and also, though not directly relevant to the present case, the union of the two link sets in their fourth positions). The result is shown in (2.31.a). Once again we add in (2.31.b) the preliminary representation of the hope content, as that representation will be the focus of the next steps.

(2.31)a.

$$
\begin{array}{|c|}
\hline
s \quad t \quad j \quad z \quad r \\
\hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
z = j \quad r = n \quad r = t \\
s\text{: } Att\left(j, \left\{ \begin{array}{l}
\left\langle BEL, \begin{array}{|c|}
\hline
t' \; s' \\
\hline
t' = n \quad t' \subseteq s' \\
s'\text{: PROG}(\hat{}e. \begin{array}{|c|}
\hline
e \\
\hline
e\text{: rain'} \\
\hline
\end{array}) \\
\hline
\end{array} \right\rangle \\
\langle HOPE, [DP] \rangle
\end{array} \right\}, \emptyset \right) \\
\hline
\end{array}
$$

b. [DP] =

$$\left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline s^2? \\ \hline s^2\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline\end{array}) \\ \hline \end{array}\ , \\ \\ \begin{array}{|c|}\hline r'?\ \rho? \\ \hline \rho'(r',t''')\text{'} \\ \hline\end{array}\ [fut] \end{array} \right\}, \begin{array}{|c|}\hline t'''\quad e\quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s^2,s^4) \\ (s^2 \supset\subset e \supset\subset s^4) \\ s^4\text{:} \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s^4 \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e\\ \hline e\text{: rain'}\\ \hline\end{array} \\ \hline\end{array} \\ \hline\end{array} \\ \hline \end{array} \right\rangle$$

There are two remaining presuppositions in the presupposition set of (2.31.b), the *stop*-presupposition and the presupposition triggered by the future tense of the complement of *hope*. The *stop*-presupposition can now be resolved by the state dref s' in the content representation of the belief constituent of (2.29), provided the durations of the state s' of the belief content and the state s'' of the presupposition can be made to fit, so that the two can be identified. (Their content descriptions match already.)

Whether the durations of the two states can be matched depends on the temporal information about s' and s'' that is contained in (2.31.a) and (2.31.b). But that information appears to give us what we need. For on the one hand (2.31.a) locates the duration of the raining state s' as lying around the time of the agent's belief state s, which holds at the utterance time because of the present tense of *believe*. And on the other hand the state s'' is located as starting after the event e which is located in the future of the utterance time. So it suffices to accommodate the assumption that the state s' lasts until the future location time t'' of the event e, an accommodation that is suggested by the presupposition of *stop* and that isn't put in doubt by any other information in representation or context.

Nothing has yet been said about the resolution of the tense presupposition in (2.31.b). In fact, there is no information in either the discourse context for the presupposition nor in the context in which the utterance of (2.6.c.1) is supposed to be taking place. In such cases accommodation of such presuppositions, to the effect that there is some future time r to which the location time stands in some temporal relation ρ, is inevitable and legitimate. This is tantamount to discarding the presupposition, so that the contribution of the future tense becomes de facto existential again.

Resolving the stop presupposition by identifying s'' with s' and discarding the tense presupposiiton leads to the Logical Form for (2.6.c.1) that is shown in (2.32).

(2.32)

$$
\begin{array}{|c|}\hline s \quad t \quad j \\ \hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
s\text{: } Att \left(j, \left\{ \begin{array}{l} \left\langle BEL, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \\ \left\langle HOPE, \begin{array}{|c|}\hline t''' \quad e \quad s' \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'',s') \\ (s'' \supset\subset e \supset\subset s') \\ s'\text{: } \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s' \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle \end{array} \right\}, \emptyset \right) \\ \hline \end{array}
$$

Since the drefs z and r have done their work, they have been eliminated from (2.32), resulting in modest simplification.

2.2.2.1 The attitudinal hierarchy

Our treatment of sentence (2.6.c.1) was based on the assumption that belief contents can serve to resolve the presuppositions triggered by the contents of other attitudes belonging to the same mental state. That seems intuitively correct. But the converse – using the content of, say, a hope content to resolve a presupposition triggered within a belief content – is not. Or at least, to the extent that such uses are acceptable at all, they seem to require some further step of adjustment. Consider the sentence in (2.33).

(2.33) John hopes that it is dry and he believes that it is going to start raining soon.

The aspectual verb *start* is the trigger of a pre-state presupposition, this time to the effect that up to the point when the transition event described by *start* takes place, a state of the kind described by the complement of *start* does not obtain. (The presupposition trigger *to stop* does the opposite. Its pre-state presupposition is that the condition expressed by its complement did hold when the transition started and asserts that after the transition it does no longer.) But can we use the content of the hope conjunct of (2.33) to resolve the *start*-presupposition of the second conjunct? Perhaps just about. But in order for this to be possible we must interpret *hope* as meaning something like 'believe with some degree of confidence and wishing that one is right about this'; and for me that is an understanding of *hope* that I can get only under duress. There may also be a second way of accepting the presuppositional connection between the first and second conjunct of (2.33), which consists in coercing the content of the belief clause into a conditional, so that the second sentence acquires the meaning 'He believes that on the assumption that his hope is true, it will start raining soon'.

The dependence asymmetry between (2.33) and (2.6.c.1) is an instance of the so-called *Attitudinal Hierarchy.*[12] Any given set of attitudinal modes – containing some subset of belief, desire, intention, planning, hope, fear, doubt, rejoicing and so on– determines a partial ordering whose members are pairs

[12]The Attitudinal Hierarchy was first introduced in the ms. part of which appeared in German translation as (Kamp 2003). But the Attitudinal Hierarchy did not make into the translation. Thanks to the participants of the Language Workshop of the Philosophy Department of UCLA in the spring of 1997, where the Attitudinal Hierarchy was extensively discussed.

<ATT$_1$,ATT$_2$> such that the content of an ATT$_2$-attribution may presuppositionally depend on the content of an ATT$_1$-attribution. If belief belongs to the set, which will be the case in pretty much any sensible application of MSDRT, then <BEL,ATT$_2$> will be part of the partial order for any member ATT$_2$ of the set (including BEL itself). That is of course pretty what can be expected even without a serious look into the matter; it is a kind of platitude that our beliefs provide the platform on which the edifice of our other attitudes is built. (That is so even for those who are a prey to wishful thinking.)[13] A further principle of apparently general validity is that or the 'positive' attitudes ATT– in the list above all are positive except for doubt – the partial ordering will include the pair <ATT,ATT>: desires can depend on other desires, intentions can depend on other intentions and so on. But in order for this principle to have some real bite, it would be good if we could give a general non-circular characterization of what it is for an attitude to be 'positive', and asa it is we have no definition at our finger tips.

For other combinations of ATT$_1$ and ATT$_2$ – ATT$_1$ $\neq$ ATT$_2$ and neither ATT$_1$ nor ATT$_2$ is belief, the question whether one can be presuppositionally dependent on the other is more delicate. One point that'll we believe holds quite general has to do with non-volitional emotively charged attitudes like hope and fear: Hopes snd fears can be referentially dependent on each other; that the ones are felt to be desirable and the others not has nothing to do with their respective doxastic strenghts. For most such pairs, it appears to be difficult to determine whether they should be members of the Hierarchy, and in individual case the verdict may depend on various factors that are hard to isolate and articulate. And in cases where the dependency relation does not seem to hold between ATT$_1$ and ATT$_2$ in general, there is the further question what coercions if any could restore the possibility of dependency under duress. Most of the work on the Attitudinal Hierarchy still has to done, in spite of the fact that the idea of such a Hierarchy has been around for more than 25 years. This is another matter that we leave for further exploration. But let us, before we leave the topic, try at least to give a flavor of the difficulties one encounters when trying to decide whether a pair <ATT$_1$,ATT$_2$> does or does not belong to the Hierarchy.

[13] In more refined attitudinal schemes, which admit beliefs of different strength (.e.g. doxastic Mode Indicators now come in the form 'BEL_d', where d is the degree to which beliefs represented with these Mode Indicators are held) Beliefs with a lower degree may referentially depend on beliefs with a higher degree, but not vice versa. Negative doxastic attitudes, such as doubts can be thought of as attitudes with a very low degree – e.g. d *and* ≤ 0 – and this can be referentially dependent on beliefs, even comparatively weak ones perhaps, but again not the other way round.

We focus on the pair $<HOPE,BEL>$. Consider the following set of three-sentences discourses.

(2.34)a. John knows that a friend of his owns a house in Cedar Park. He **believes** that she has rented it out to some lawyer. He **hopes** the lawyer **takes** good care of it.

b. John knows that a friend of his owns a house in Cedar Park. He **hopes** that she has rented it out to some lawyer. He **believes** the lawyer **takes** good care of it.

c. John knows that a friend of his owns a house in Cedar Park. He **hopes** that she has rented it out to some lawyer. He **believes** the lawyer **will take** good care of it.

d. John knows that a friend of his owns a house in Cedar Park. He **hopes** that she has rented it out to some lawyer. He **believes** the lawyer **would take** good care of it.

e. John knows that a friend of his owns a house in Cedar Park. He **hopes** that she has rented it out to some lawyer. He **believes** **that if she has,** the lawyer **will take** good care of it.

In all these discourses the matter at issue is the pair $<\mathrm{ATT}_1,\mathrm{ATT}_2>$ where ATT_1 is the Mode Indicator of the second sentence and ATT_2 the Mode Indicator of the third. For the first discourse, in (2.34.a), this is the pair $<BEL,HOPE>$. This pair does belong to the Attitudinal Hierarchy according to what we said above. So the referential dependence of the third sentence on the second – because of the definite DP *the lawyer*, for which the second sentence should provide the resolution – is unproblematic in this case, which is an agreement with the intuition that this discourse is perfectly felicitous.

For all the other discourses in (2.34) the corresponding Indicator pair is $<HOPE,BEL>$, and it is these discourses that matter for our question about this pair. Here are our judgments (for what they are worth):

(2.34.b) and (2.34.c) are both not good. ((2.34.c) may be marginally better then (2.34.b), but the difference is marginal for us.)

(2.34.d) and (2.34.e) are both felicitous.

The first moral that we infer from these somewhat tentative judgments is:

$<HOPE,BEL>$ is not a member of the Attitudinal Hierarchy.

Secondly, that both (2.34.d) and (2.34.e) are felicitous is easily explained. In (2.34.e) the third sentence has the form of a belief whose content is a conditional. The antecedent of this conditional provides the resolution of the definite description *the lawyer* in its consequent, according to generally accepted principles of presupposition resolution (Karttunen 1974) and much further literature that has followed Karttunen's lead. So the third sentence of (2.34.e) doesn't put the pair $<HOPE,BEL>$. The story about (2.34.d) is much the same, given the property of counterfactual *would* that it presupposes an antecedent and that when no such antecedent is present overtly, it is usually easy to accommodate one. (Frank 1997), (Frank & Kamp 1997). In (2.34.d) this accommodated antecedent would be the proposition that John's friend has rented out her house to a lawyer, that accommodated antecedent resolves the description *the lawyer* and once more there is no direct test for our question about $<HOPE,BEL>$.

As the last two discourses in (2.34) aren't directly relevant to our test, you might wonder why we have included them. The point is this. In (2.34.b) and (2.34.c) the last sentence isn't a good follow-up to the second sentence, we observed. but perhaps these discourses aren't completely out. And if speakers do not discard them as categorically wrong, then that is, we conjecture, because one can sort of coerce the third sentence into a conditional belief, of the form: 'if my hope is true then the lawyer take good care of the house'. Such a coercion isn't properly licensed by English grammar, which is why the discourses in (2.34.b) and (2.34.c) aren't very felicitous. But if one is confronted with such a discourse, and has to do something with it, then the best one can do is to apply this not quite legitimate coercion.

We do not know this for certain, but we suspect that this is the way around violations of referential dependency constraints more generally: when you are confronted with a sentence pair $<\mathrm{S}_1,\mathrm{S}_2>$ where both S_1 and S_2 are sentences with Mode Indicators, A_1 and A_2 and S_2 is referentially dependent on S_1 and $<\mathrm{ATT}_1,\mathrm{ATT}_2>$ does not belong to the Attitudinal Hierarchy, then reinterpret S_2 as of the form $\mathrm{ATT}_2(p \rightarrow q)$, where p is the content in the scope of ATT_1 in S_1 and q the content in the scope of ATT_2 in S_2.

If this hypothesis is correct, then many prima facie violations of Attitudinal Hierarchy-related violations of referential dependence will give rise to MSDs in which the violating sentence has been reinterpreted as involving a conditional. (2.35) exemplifies this for the discourse in (2.34.b).

$$(2.35)\left\{ \begin{array}{l} \left\langle KNOW, \begin{array}{|c|}\hline f\quad h \\ \hline \text{friend-of'}(f,i)\qquad \text{house'}(h) \\ \text{own'}(f,h) \\ \\ \hline \end{array} \right\rangle \\ \\ \left\langle HOPE, \begin{array}{|c|}\hline l \\ \hline \text{lawyer'}(l)\qquad \text{rent'}(f,h,l) \\ \hline \end{array} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline \\ \hline \begin{array}{|c|}\hline l' \\ \hline \text{lawyer'}(l')\qquad \text{rent'}(f,h,l') \\ \hline \end{array} \Rightarrow \begin{array}{|c|}\hline \\ \hline \text{take-good-care'}(l',h) \\ \hline \end{array} \\ \hline \end{array} \right\rangle \end{array} \right\}\text{[14]}$$

But this has been no more than a tentative step toward trying to become clearer on just one potential member of the Attitudinal Hierarchy. Nan that is just one tiny spot from a vast territory. For a language like English the territory is as vast as is it is because the language has countless attitudinal verbs. Here is a list of nothing but doxastic verbs, produced by a brief reflection in the course of writing this: *believe*, *know*, *assume*, *suppose*, *conjecture*, *hypothesize*, *speculate*, *suspect*, *expect*, *be convinced*, *be persuaded*, *be certain*. And far from all attitudinal verbs are doxatic. Besides the verbs *desire*, *want*, *hope* and *fear*, which have been mentioned at one point or another in this text so far, there are quite others. And when the total number of those verbs is N, then the total number of candidates for the Attitudinal Hierarchy is N^2. Much work that is still to be done.[15]

[14] A warning about the Mode Indicator 'KNOW' that is used in (4.7). The attitudinal mode it indicates is a purely doxastic one, a mode like belief, but involving a higher degree of certainty. The difference between the modes indicated by 'KNOW' and 'BEL' corresponds roughly to first person uses of the verbs *know* and *believe*, which agents use to express their own attitudes. 'KNOW' is not supposed to differ from 'BEL' in the sense of involving an external judgment that the attitude ascribed to the agent is towards a content that is in fact true (and that the attitude is 'properly grounded' **(ref)**, an aspect of the meaning of *know* that tends to be present when it is used in the third person, in attributions to others.
We have also simplified this MSD in that all temporal information has been left out from the content representations of its Propositional Attitudes

[15] One important factor among those that determine whether or not a pair $<ATT_1,ATT_2>$ belongs to the Attitudinal Hierarchy, when both ATT_1 and ATT_2 are Mode Indicators expressed by doxastic/epistemic verbs is the degree of confidence these verns express. (Degen & Tonhauser 2022) is a large scale corpus-based study of such

We will briefly return to the Attitudinal Hierarchy in Section 2.3 of Chapter 4.

2.2.2.2 Back to sentence (2.6.c.2): Covering the same ground, but lower down in the syntax

Here, once more, is sentence (2.6.c.2) in full:

(2.6.c.2) John believes that it is raining and hopes that it will stop raining.

The point of this sentence, after we have already dealt with its close analogue (2.6.c.1) in Section 3, is that the connection between belief and hope can be established lower down than at the level of complete clauses. Intuitively there is nothing remarkable about this. But given the general assumptions we have made about how semantic representations are computed, in sentences like (2.6.c.2) the connection between the two attitude attributions has to be established at a lower level than the one at which this becomes possible, and therefore has to be done, in the DRS construction for a sentence like (2.6.c.1). The difference is not a big one. But it is worth our while to see the details for cases like (2.6.c.2) as well as for (2.6.c.1), to see exactly what the difference is: merging *Att*-predications is an operation that can occur at different levels of sentence syntax.

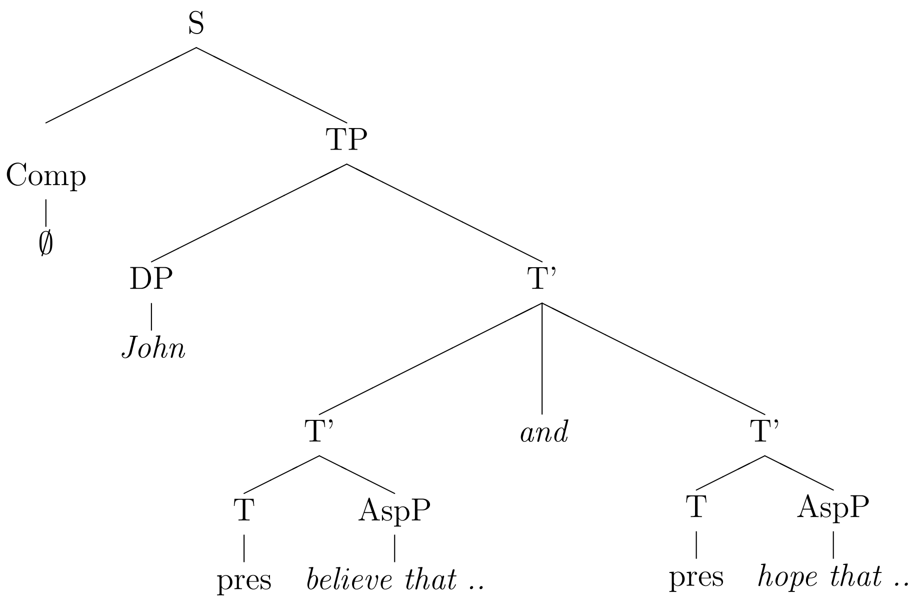
The first matter to be decided about (2.6.c.2) is the syntactic level at which the conjunction is formed. The most plausible assumption about this is, as far as we can see, that the conjunction is formed at the level of T' – that is, after the two verbs have obtained their respective tenses (which in the case of (2.6.c.2) are both the simple present), so this is the one we will adopt for the following discussion.[16] I will assume that the relevant part of the syntactic tree for (2.6.c.2) has the form shown in (2.39).

---

and other attitude verbs that explores their degrees of *factivity*. Factivity isn't the same thing as strength of conviction. But the two notions are not very unlike each other; and the method used in this large scale exploration might well be applicable to detrees of conviction, and perhaps to other factors that are relevant to the Attitudinal Hierarchy.

[16] In principle there would appear to be other possibilities too, e.g. the one according to which the conjunction is formed at the AspP level and that the 'pres' feature of the T-node for the syntactic structure with this conjunctive AspP then has the morphological effect of 'spreading over' the two verbs. But I cannot detect much plausibility in this alternative, or any other I can think of. And in any case the problem we have to address will arise for any of these possibilities.

(2.36)



Representations for the two T' conjunctions have, for all practical purposes, already been computed

(2.37)

a. $$\left\langle \left\{ \begin{array}{|c|} \hline r_1? \quad \rho_1? \\ \hline \rho_1(r_1,t_1) \\ \hline \end{array} \right\}, \begin{array}{|c|} \hline s_1 \\ \hline t_1 = n \quad t_1 \subseteq s_1 \\ s_1\text{: } \mathit{Att}\left(\underline{x}, \left\{ \langle \mathit{BEL}, [\mathrm{DP}_1] \rangle \right\}, \emptyset \right) \\ \hline \end{array} \right\rangle$$

b. $$\left\langle \left\{ \begin{array}{|c|} \hline r_2 \quad \rho_2? \\ \hline \rho_2(r_2,t_2) \\ \hline \end{array} \right\}, \begin{array}{|c|} \hline s_2 \\ \hline t_2 = n \quad t_2 \subseteq s_2 \\ s_2\text{: } \mathit{Att}\left(\underline{x}, \left\{ \langle \mathit{HOPE}, [\mathrm{DP}_2] \rangle \right\}, \emptyset \right) \\ \hline \end{array} \right\rangle$$

(The underlined argument $\underline{x}$ is not a dref but a 'slot marker', which at some point will have to be replaced by a dref that properly fills the argument slot it marks – in this case by the dref representing the denotation of the grammatical subject.

The representation of the conjunction of these two T' representations is simply their merge : the union of the two presupposition sets followed by the merge of the two DRSs:

(2.38)

$$\left\langle \left\{ \begin{array}{c} \boxed{\begin{array}{c} r_1? \quad \rho_1? \\ \hline \rho_1(r_1,t_1) \end{array}} \\ \\ \boxed{\begin{array}{c} r_2? \quad \rho_2? \\ \hline \rho_2?(r_2,t_2) \end{array}} \end{array} \right\}, \boxed{\begin{array}{c} s_1 \quad s_2 \\ \hline t_1 = n \quad t_1 \subseteq s_1 \\ t_2 = n \quad t_2 \subseteq s_2 \\ \\ s_1\colon \mathit{Att}\left(\underline{x}, \left\{ \langle \mathit{BEL}, [\mathrm{DP}_1]\rangle \right\}, \emptyset\right) \\ s_2\colon \mathit{Att}\left(\underline{x}, \left\{ \langle \mathit{HOPE}, [\mathrm{DP}_2]\rangle \right\}, \emptyset\right) \end{array}} \right\rangle$$

We are now in much the same position that we were in when we constructed the representation for (2.6.c.1): the states described by the two *Att*-predications are located at the same time – they are both located at n – and these states ascribe mental states to the same agent (the one who is going to fill the slots identified by the slot marker at the next level when this T' representation is combined with the subject DP *John*. So the conditions are satisfied for merging the two *Att*-state descriptions into one, with the same effect as in our representation construction for (2.6.c.1). The result of performing this operation right here is given in (2.39). The remainder of the story – how the *stop*-presupposition in the hope part can be resolved using the content representation of the belief – is possible just as it was in the construction for (2.6.c.1). Note also that the merge of the two *Att*-predications is possible in this case even before resolutions of the tense presuppositions. This is because we have opted for the deictic treatment of the two present tenses in the *believe* and the *hope* clause, which sets both location times equal to the utterance time. The tense presupposition plays no role of significance in this case, its only possible resolution is the one that sets r equal to n and ρ to $=$.

$$(2.39)\left\langle \left\{ \begin{array}{c} \boxed{\begin{array}{c} r_1? \quad \rho_1? \\ \hline \rho_1(r_1,t_1) \end{array}} \\ \\ \boxed{\begin{array}{c} r_2? \quad \rho_2? \\ \hline \rho_2(r_2,t_2) \end{array}} \end{array} \right\}, \boxed{\begin{array}{c} s \\ \hline t_1 = n \quad t_1 \subseteq s \\ t_2 = n \quad t_2 \subseteq s \\ \\ s\colon \mathit{Att}\left(\underline{x}, \left\{ \begin{array}{c} \langle \mathit{BEL}, [\mathrm{DP}_1]\rangle \\ \\ \langle \mathit{HOPE}, [\mathrm{DP}_2]\rangle \end{array} \right\}, \emptyset\right) \end{array}} \right\rangle$$

2.2.3 Attitude reports with pronouns and indefinite antecedents

The next sentence from our list in (2.6) is (2.6.d):

(2.6.d) John is convinced that there is a mistake in his proof and he is dead set on finding it.

This is the type of sentence that was the immediate reason for developing MSDRT. In a sense the pronoun *it* in (2.6.d) is an example of a 'donkey pronoun': it has an indefinite antecedent, but one that occurs in a different sentence or clause, and in such a way that the principles of variable binding in Predicate Logic, the Lambda Calculus and Montague Grammar can't be applied to them in the way one would have expected. But as argued in the introduction, the pronoun in (2.6.d) cannot be dealt with adequately either with the method of classical DRT, since that method foregoes assigning a separate propositional content to the clause containing the pronoun. A propositional content gets assigned only to the update of the content of the DRS for the part that contains the pronoun's indefinite antecedent with the DRS for the part containing the pronoun, but not for the latter DRS on its own; and that isn't the right result for sentences like (2.6.d), where antecedent and pronoun occur in the complements of attitude verbs of different attitudinal mode. (It was for sentences of this type that MSDRT was originally designed and it is also the semantics of such sentences that complicates its model theory, as we will see later.)

So much for a recapitulation of why a sentence like (2.6.d) really should appear somewhere in our first list of DRS construction tasks in MSDRT. But after the examples we have dealt with at this point there isn't much more that needs to be said. In fact, (2.6.d) is very similar to (2.6.c.1) in that both consist of a non-doxastic (or not purely doxastic) attitude attribution whose content depends on the content of the attribution of a doxastic attitude. The only difference is that in (2.6.c.1) the dependence involved the presupposition of the verb *to stop*, whereas in (2.6.d) it is the anaphoric dependence of *it* on the indefinite *a mistake in his proof*. And even that isn't a very big difference, once one adopts Van Der Sandt's insight that 'propositional' presuppositions like that of *stop* and nominal anaphora involving pronouns and other types of noun phrases are two sides of what is but a slim coin. In MSDRT, in which presupposition and anaphora are handled as in DRT (as outlined in (Kamp et al. 2011)[17]), anaphoric expressions trigger 'anaphoric presuppositions', the resolution of which consists in finding an antecedent for the expression. That a presupposition is anaphoric in this sense – i.e. that drefs have to be found in the representation context that can serve as antecedents for one or more of the drefs occurring in the representation of the presupposition – is indicated by providing the 'anaphoric' dref with a question mark. (We have already seen this for the presuppositions triggered by tense features.)

The only other respect in which (2.6.d) differs from the earlier examples in (2.6) is the choice of their 'matrix verbs', *be convinced* and *be dead set on*. I chose those expressions after groping about for attitude verbs that would sound natural in the context of the content I wanted this sentence to express and these were the ones that came to mind. How happy a choice this is may be a matter for debate. But the general point is that in their actual practices of attitude attribution speakers make use of a wide and open-ended repertoire, with subtle distinctions between the different items that can be found there. (If there were no differences between these expressions, there could hardly be any point in searching the repertoire for the right item that would best fit one's communicative intentions. We will encounter many examples of this when applying MSDRT to the vocabulary that English uses for describing intentions, plans and actions.).

One of the tasks for a fine-grained semantics of the attitudes is an account of precisely what the meanings of these different items are and how they differ from each other. The general issue here is one that I already touched upon in the little I said about *hope* as distinct from *believe*, *want* and *wish*. At that point I delegated a proper account of such distinctions to later, and here I won't do more than just repeat that. In our summary treatment of the DRS construction for (2.6.d) I will simplify this aspect of the sentence by identifying *be convinced* with *believe* and *be dead set on* with *intend*, an attitudinal verb with a standardized representation in MSDRT in the form of the Mode Indicator *INT*.

The first conjunct of (2.6.d) contains the pronoun *his*. Possessive pronouns are analyzed in DRT[18] (and MSDRT) as compound expressions consisting of (i) a personal pronoun and (ii) a 'possession' component, which relates the referent of the pronoun to that of the DP that contains it as determiner.

[17]See also the details in (Kamp 2021*a*), Ch. 4.

[18](Kamp 2021*a*), Ch. 4

This possession relation can take many more specific forms, from ownership, control or authorship to relations that depend heavily on special contexts. (For instance, in the case of (2.6.d), and assuming, as we will, that the pronoun is anaphoric to the subject *John*, the relation expressed by *his* is perhaps most easily understood as that of the proof being John's own (in the sense that he was the one who found it); but in the right context it could also be the one of several alternative proofs of some well-known theorem that John has been assigned to present at the next session of a graduate seminar he is participating in; and so on.[19] Here I will assume without further discussion that the proof spoken of in (2.6.d) is one that John came up with. But in the Logical Form we will construct the possession relation is nevertheless represented by the portmanteau possession predicate *OF*. Taking this and the resolution of *his* for granted, and proceeding as for several earlier present tense belief attributions, we obtain for the first conjunct of (2.6.d) the DRS in (2.40).

A second point about the representation in (2.40) and its construction, which also has nothing to do with propositional attitudes as such, concerns the *there is* construction of (2.6.d). I have simplified the matter by simply putting the more controversial issues surrounding this construction aside. (2.40) adopts the widely accepted assumption that the *there is* construction expresses the existence of an entity that satisfies the complex noun phrase – here: *a mistake in his proof* – which follows and complements the verb *be*. In DRT this assumption is implemented by letting the construction create its own sub-DRS, which contributes to the DRS K in whose Condition Set it occurs[20] the existence of entities represented in its Universe, with the properties asserted of them in the Condition Set of this sub-DRS. As part of representing the *there be* construction a fresh dref is introduced into the Universe of the sub-DRS – in the present example this is the dref m – and the claim that m satisfies the predicate expressed by the noun phrase following the *be* of *there be* is represented in the sub-DRS's Condition Set. (2.40) shows the result for the belief conjunct of (2.6.d).

[19]For some salient references on the topic of possession relation and their expression by possessive constructions see (Barker 1995), (Partee & Borschev 2003).

[20]A DRS that occurs in the Condition Set of a larger DRS thereby treated as a DRS Condition. It is like a DRS Condition which is formed from a single DRS (like DRT's Negation Conditions) but with an invisible and semantically otiose operator. That is, for such a DRS Condition K' to be verified by an embedding f it is necessary and sufficient that f can be extended to an embedding g which is also defined for the Universee of K' and which verifies all Conditions in the Condition Set of K'.

(2.40)

$$
\begin{array}{|c|}
\hline
s \quad t \quad j \quad x \quad r \\
\hline
\begin{array}{c}
t = n \quad t \subseteq s \quad \text{John'}(j) \\
x = j \quad r = n \quad r = t \\
\\
s\text{: } Att \left(j, \left\{ \left\langle BEL, \begin{array}{|c|}
\hline
t' \quad s' \quad p \\
\hline
t' = n \quad t' \subseteq s' \\
\text{proof'}(p) \quad \text{OF}(p,x) \\
\\
s'\text{: } \begin{array}{|c|}
\hline
m \\
\hline
\text{mistake-in'}(m,p) \\
\hline
\end{array} \\
\hline
\end{array} \right\rangle \right\}, \emptyset \right)
\end{array} \\
\hline
\end{array}
$$

Although in the belief content DRS of (2.40) m is in a subordinate DRS, it is nonetheless accessible to anaphoric pronouns as if it were a member of the Universe of that DRS. (This is a fact about the entities whose existence is asserted in *there be* sentences. This may seem plausible enough, but it isn't, I think, more than that. A language is imaginable with a construction that is much like the English *there be* construction, but where the entity whose existence is asserted by it, is not accessible to the pronouns of that language. But plausible or not, this is a fact about English *there be*. Its formal implication is that the sub-DRS may be merged with the one in whose Condition Set it occurs. This doesn't alter the truth conditions of the larger DRS and it gets the drefs of the sub-DRS in the right place. The result is shown in (2.41).

(2.41)

$$
\begin{array}{|c|}
\hline
s \quad t \quad j \quad x \quad r \\
\hline
\begin{array}{c}
t = n \quad t \subseteq s \quad \text{John'}(j) \\
x = j \quad r = n \quad r = t \\
\\
s\text{: } Att \left(j, \left\{ \left\langle BEL, \begin{array}{|c|}
\hline
t' \quad s' \quad p \quad m \\
\hline
t' = n \quad t' \subseteq s' \\
\text{proof'}(p) \quad \text{OF}(p,x) \\
s'\text{: mistake-in'}(m,p) \\
\hline
\end{array} \right\rangle \right\}, \emptyset \right)
\end{array} \\
\hline
\end{array}
$$

To construct a representation for the second conjunct of (2.6.d) we assume that the 'attitude verb' *be dead set on* can be represented by 'INT', the Mode Indicator for intention; and we pass over the presupposition triggered by the present tense of *is*, assuming that tense processing leads to the same contributions that are shown in (2.40) for the first conjunct. We'll also assume that *he* is interpreted as anaphoric to *John*, saving ourselves the trouble of explicitly representing the anaphoric presupposition triggered by *he* and then explicitly resolving that representation. On the other hand the presupposition that is directly related to what matters about this sentence, the one triggered by the pronoun *it*, is displayed explicitly in the preliminary representation for this conjunct, shown in (2.42).

(2.42)

$$\left\langle \left\{ \begin{array}{|c|}\hline z? \\ \hline \text{non-human}(z) \\ \hline \end{array} \right\}, \begin{array}{|c|} \hline s' \quad t' \\ \hline t' = n \quad t' \subseteq s' \\ s'\colon Att\left(j, \left\{ \left\langle INT, \begin{array}{|c|}\hline e \\ \hline t' \prec e \\ e\colon \text{find'}(i,z) \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\ \hline \end{array} \right\rangle$$

Intuitively it is clear that *it* is anaphoric to the indefinite *a mistake in his proof* and that this must therefore be the result of resolving the presupposition in (2.42). This means that z must be identified with the dref m that represents this noun phrase. But how can this step be justified within the system of construction principles we are in the process of developing?

The first step is the same as in our construction for (2.6.c): Since the two *Att*-predications in (2.41) and (2.42) have the same attributee and are located at the same time, they are descriptions of one and the same mental state and thus can be merged into a representation that makes this explicit.

$$
(2.43)\quad
\begin{array}{|c|}
\hline
s \quad t \quad j \quad x \quad r \\
\hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
x = j \quad r = n \quad r = t \\
s\text{: } Att \left(j, \left\{ \begin{array}{c}
\left\langle BEL, \begin{array}{|c|}
\hline
t' \quad s' \quad p \quad m \\
\hline
t' = n \quad t' \subseteq s' \\
\text{proof'}(p) \quad \text{OF}(p,j) \\
s'\text{: mistake-in'}(m,p) \\
\hline
\end{array} \right\rangle \\[2ex]
\left\langle \left\{ \begin{array}{|c|}
\hline
z? \\
\hline
\text{non-human}(z) \\
\hline
\end{array} \right\}, \left\langle INT, \begin{array}{|c|}
\hline
e \\
\hline
t' \prec e \\
e\text{: find'}(i,z) \\
\hline
\end{array} \right\rangle \right\rangle
\end{array} \right\}, \emptyset \right) \\
\hline
\end{array}
$$

We want to resolve the presupposition of (2.43) by identifying z with m. But is that legitimate? The answer is yes and follows from the Attitudinal Hierarchy, according to which the content representations of beliefs are available as discourse contexts for presupposition resolution of presuppositions that are part of the content representations of Propositional Attitudes of arbitrary Attitudinal Mode that are part of the same MSD. The dref m in the Universe of the belief content representation in (2.43) is therefore accessible to the position of z?. Identifying z with m gives us the Logical Form of (2.6.d), shown in (2.44).

$$
(2.44)\quad
\begin{array}{|c|}
\hline
s \quad t \quad j \quad x \quad r \\
\hline
t = n \quad t \subseteq s \quad \text{John'}(j) \\
x = j \quad r = n \quad r = t \\
s\text{: } Att \left(j, \left\{ \begin{array}{c}
\left\langle BEL, \begin{array}{|c|}
\hline
t' \quad s' \quad p \quad m \\
\hline
t' = n \quad t' \subseteq s' \\
\text{proof'}(p) \quad \text{OF}(p,x) \\
s'\text{: mistake-in'}(m,p) \\
\hline
\end{array} \right\rangle \\[2ex]
\left\langle INT, \begin{array}{|c|}
\hline
e \quad z \\
\hline
t' \prec e \quad z = m \\
e\text{: find'}(i,z) \\
\hline
\end{array} \right\rangle
\end{array} \right\}, \emptyset \right) \\
\hline
\end{array}
$$

2.2.4 Proper Names in Attitude Reports. The Communication-Theoretic Approach to Formal Semantics

The main topic of this section is sentence (2.6.e), repeated below.

(2.6.e) John believes that Mary is in Paris.

This sentence may look as innocent as any on its list. But the issue that it will require us to address differs dramatically from anything we have said so far in Section 2.3. The treatment of proper names in MSDRT involves a radically different view of how natural language semantics can and sometimes should be done. This alternative approach can be described as *communication-theoretic*. It takes the form of analyzing how the interpretations that recipients of spoken utterances or written texts construct for those texts or utterances change their mental states. In MSDRT this approach is implemented by using MSDs to characterize those mental states. Crucial is that for many aspects of language it isn't only the mental state that results from interpretation, but also the mental state the interpreter is in when interpretation starts, since that mental state is among the contextual resources he may need for interpreting the linguistic input. The way speakers handle proper names is one example of this.

As a first illustration of this alternative way of doing semantics we start with an example lifted from(Kamp 2022*a*). (What follows is from Section 3.5 of that paper.)

In this section we look in some detail at how according to MSDRT utterance production and interpretation are related to the mental states of source and interpreter. We will be looking at just one sentence, an utterance made by a speaker S to a hearer H.

(2.45) S: I read a short story by Gogol.

We consider two scenarios in which S utters and H interprets (2.45). In the first H doesn't have an ER for Gogol, in the second he does. (2.46) gives the MSD for the relevant part of the mental state of S, assuming that S herself believes the information that is expressed in (2.45) and that she has a *Gogol*-labeled ER for Gogol that puts her in a position to use the name *Gogol* in (2.45). (2.47) describes the relevant part of H's mental state in the first of our two scenarios just before his interpretation of (2.45) gets under way. Note that the only ER shown in (2.47) is an ER for the speaker S. The fact that (2.47) does not display an ER for Gogol is to be understood in this case as an indication that the mental state partially described by (2.47) doesn't have such an ER.

$$(2.46)\left\{ \begin{array}{l} \left\langle [ENT, g_S]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(g_S) \\ \text{Named}(g_S, Gogol) \\ \hline \end{array}\ , \mathcal{K}_{Gogol} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline e \;\; y \\ \hline e \prec n \quad \text{short-story'}(y) \quad \text{by'}(y,g_S) \\ e\text{: read'}(i,y) \\ \\ \hline \end{array} \right\rangle \end{array} \right\}$$ [21]

$$(2.47)\left\{ \left\langle [ENT, s_H]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\ , \mathcal{K}_S \right\rangle \right\}$$

When H is in the mental state described in (2.47), then his interpretation of (2.45) requires the accommodation of an ER for Gogol. The result of this accommodation and the construction of H's representation of the content of (2.45), which, we assume, H accepts as true as well and thus adopts as a belief, leads to the mental state described in (2.48).

[21] Notation: 1. The symbol 'g_S' that has been chosen here for the distinguished dref of S's ER for Gogol is motivated by wanting distinct symbols for this dref and the distinguished dref of H's ER for Gogol in (2.49) below. (The subscripts $_S$ and $_H$ are chosen purely for mnemonic reasons – $_S$ means that we are dealing with a dref in an MSD for S, and $_H$ likewise that we are dealing with a dref in an MSD for H. They have no significance for the theory.) 2. The content DRS of the belief in (2.46) belongs to a DRS language that extends beyond the description of DRS languages in Section 2 of Ch. 1. As noted in Section 2.1, the original motivation for DRT was to account for certain properties of tense forms in French and English. Versions of DRT in which these motivating ideas received an early implementation can be found in (Kamp 1981*a*) – see also the English version (Kamp 2017) – and in (Kamp & Reyle 1993). The intuitive meaning of the DRS in (2.46) is that there was an event e of S reading some short story by Gogol at some time in the past of the time n at which S is in the mental state that (2.46) describes.

$$
(2.48)\left\{
\begin{array}{l}
\left\langle [ENT, s_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person'}(s_H) \\ \hline \end{array}\ ,\ \mathcal{K}_S \right\rangle \\[2em]
\left\langle [ENT, g_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person'}(g_H) \\ \text{Named}(g_N, \textit{Gogol}) \\ \hline \end{array}\ ,\ \left\{ \begin{array}{|c|}\hline e_1 \\ \hline e_1 \prec n \\ e_1\text{: ref}(s_H, \textit{Gogol}, g_H) \\ \hline \end{array} \right\} \right\rangle \\[2em]
\left\langle BEL, \begin{array}{|c|}\hline e_2\quad y \\ \hline e_2 \prec n \quad \text{short-story'}(y) \quad \text{by'}(y, g_H) \\ e_2\text{: read'}(s_H, y) \\ \hline \end{array} \right\rangle
\end{array}
\right\}
$$

New in (2.48) is the *vicarious anchor* of H's accommodated ER for Gogol. Vicarious anchors are like perceptual anchors testimonies to causal links that connect ERs with the entities they represent. But they differ from perceptual anchors in that the relations to which they testify are not perceptual relations between the agent and the represented entity, but rather relations between the agent and some reference made to the entity made by some other agent. In the case we are considering the reference has been made by the speaker S. The vicarious anchor that makes up the anchor set of the ER for Gogol in (2.48) records that it was the reference to Gogol by S (by means of the proper name *Gogol*) that gave rise to this ER and that makes it into a representation of Gogol. (This is the meaning of the Condition 'e_1: ref(s_H,*Gogol*,g_H)'.)

Vicarious anchors for entity representations play a crucially important part in applications of MSDRT to the analysis of reference by definite noun phrases, and especially to MSDRT's account of names.

The general form we adopt for vicarious anchors is schematically presented in (2.49).

$$
(2.49)\quad \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}(\sigma, \gamma, \alpha) \\ \hline \end{array}\ ,
$$

where σ is the distinguished dref for the interpreter's ER for the speaker, γ is the referring expression she has used and α the distinguished dref of the ER of which the interpreter assumes that the speaker has used γ to refer to its referent.

In (2.49) e is the act of reference that gives rise to H's accommodated ER, s is the source of the reference (the speaker or text), γ the referring expression used by s and α the distinguished dref of the ER of which (2.49) is an anchor. (Thus in the vicarious anchor of the ER for Gogol in (2.48), s is the dref s_H, γ is the expression *Gogol* and α is the distinguished dref g_H.)

We now consider the case where H does have a *Gogol*-labeled ER for Gogol before S says (2.45) to him. In this case H's mental state before he starts his interpretation of (2.45) can be described as in (2.50).

$$(2.50)\left\{ \begin{array}{l} \left\langle [ENT, s_H]\,,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\,,\ \mathcal{K}_S \right\rangle \\ \\ \left\langle [ENT, g_H]\,,\ \begin{array}{|c|}\hline \\ \hline \text{person}(g_H) \\ \text{Named}(g_H, \textit{Gogol}) \\ \hline \end{array}\,,\ \mathcal{K}_{Gogol} \right\rangle \end{array} \right\}$$

According to MSDRT the result of processing (2.45) is the one shown in (2.51).

$$(2.51)\left\{ \begin{array}{l} \left\langle [ENT, s_H]\,,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\,,\ \mathcal{K}_S \right\rangle \\ \\ \left\langle [ENT, g_H]\,,\ \begin{array}{|c|}\hline \\ \hline \text{p'n}(g_H) \\ \text{N'd}(g_H, \textit{Gogol}) \\ \hline \end{array}\,,\ \mathcal{K}_{Gogol} \cup \left\{ \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}(s_H, \textit{Gogol}, g_H) \\ \hline \end{array} \right\} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline e \quad y \\ \hline e \prec n \quad \text{short-story'}(y) \quad \text{by'}(y, g_H) \\ e\text{: read'}(s_H, y) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

Compare the anchor set of the Gogol ER in (2.51) with that the Gogol ER in (2.48). The vicarious anchor that is explicitly displayed is a member of both sets. But in the Gogol ER of (2.48) it is the only anchor whereas in the anchor set of (2.51) it will be one of more. The *Gogol*-labeled ER that H has before S's utterance reaches him will already have a non-empty anchor set, with anchors witnessing his earlier causal relations with the referent. The assumption here is that even if the vicarious anchor that witnesses his current contact with the referent, via S's use of *Gogol*, isn't necessary to provide the ER with a referent – that has already been done by the anchors in $\mathcal{K}_{Gogol}$ – H's interpretation of S's use of *Gogol* via his *Gogol*-labeled ER will nevertheless leave this vicarious anchor as a trace of this interpretation event.[22]

(2.52) summarizes MSDRT's account of the 'standard use of proper names' (Kamp 2015), of which S's use of *Gogol* in (2.45) is one example.

(2.52)(i) A speaker or author can make a legitimate standard use of a proper name N only when she has an N-labeled ER and intends to use N to refer to the entity represented by this ER.

(ii) The interpreter of a speaker's or author's standard use of N must either have an N-labeled ER that he can assume to be coreferential with the ER the speaker or author has used; otherwise he must accommodate by forming a new N-labeled ER that is coreferential with the speaker's or author's use of N by stipulation, as indicated by its vicarious anchor.

(iii) The communication between producer and interpreter mediated by their production and interpretation of N is *successful* only if the ER used by the interpreter is coreferential with the ER used by the producer.

The story that has just been told about the interpretation of (2.45) is a story about how interpreters deal (or should deal) with names in utterances that come their way and that they need to make sense of, as part of all the other things interpreters have to do in accordance with the rules of their language. But what does this tell us about the *truth conditions* of (2.45)?

For an answer to this question we have to turn to model theory, eventually to the model theory for the DRS languages of MSDRT, which we have not yet said anything about. But for a sentence like (2.45) and the change that its interpretation produces in the mind of a language-abiding interpreter H

[22] Note that when H uses a wrong ER in his interpretation of S's use of *Gogol* – one that does not represent the same individual as the one that S's used in her use of *Gogol* – then this does not only causes a misalignment between him and S, but also an impairment of his own ER, which now has an anchor that points to another referent than the anchors in the anchor set of his ER. This is only one way in which anchor sets can get corrupted. The question of anchor and anchor set corruption is a theme in its own right, which we will not pursue any further here.

only the model theory for DRS languages of DRT is required (which we have talked about in detail).

The answer comes to this. The crucial addition to H's mental state because of his interpretation of S's utterance of (2.45) is the belief constituent of (2.51). This constituent has a content representation in the form of a DRS, to which I will refer as 'K_{bel}' in the following lines. This is the Logical Form that H assigns to S's utterance of (2.45). If H proceeded in agreement with the interpretation rules of his language, then K_{bel} can be regarded as the Logical Form that a Logical Form approach like DRT would assign to an utterance of (2.45). So the truth conditions for our treatment of (2.45) predicts for utterances of this sentence are those that the model theory of the relevant DRS language (one that counts K_{bel} among its well-formed formulas) defines for K_{bel}.

But let me, with an eye on things that will come up later, say a little more about this. The models for DRS languages we are using are IH models (Intensional History models) families of extensional models indexed by world time pairs $<w,t>$. Consider all those indices $<w,t>$ where some speaker S utters (2.45) and a recipient H interprets that utterance in accordance with the rules of the language and thereby arrives at a new belief with a content representation K that will be like K_{bel} in (2.51). And suppose that H uses for his interpretation of *Gogol* in his interpretation some entity representation that is anchored to the individual **g**[23] the model. Then H's belief will be true at $<w,t>$ iff there is a verifying embedding f of K_{bel} which maps g_H to **g**.[24]

Note also that when truth evaluation of Logical Forms of sentences in IH models is restricted in the way just described, then our earlier treatment of proper names, involving a dref to represent the referent x_N of a name N together with the predication 'N'(x_N)', will do just as well: Evaluation is to occur only at indices $<w,t>$ at which the represented sentence or discourse is produced and interpreted by some recipient H, who interprets the names

[23] Recall the convention about the use of bold face letters adopted in Ch. 1. (See footnote 13 on page 46.)

[24] An alternative way of telling this story assumes that the speaker S is among the interpreters of her own utterance. (Not implausibly, I suggest; we do have a sense of what it is we are saying when we make statements, and this sense rests of what we can recognize them to mean. This is so also in those cases where we do not really know in full detail what we are going to say until the words are out. Language producers are, ipso facto almost, self-interpreters. And perhaps they should be given preference to other interpreters of their utterances, when we want to determine the content of their utterances in terms the mental state changes that utterance interpretations produce.

in it as referring to entities for which he has anchored representations. Admissible embeddings are then those that map x_N onto the entity that H has assigned to the given occurrence(s) of N in this way.

After this interlude about (2.45) we turn at last to (2.6.e), repeated here once more.

(2.6.e) John believes that Mary is in Paris.

Common between (2.6.e) and (2.45) is that both have occurrences of proper names. but of course that is something that (2.6.e) doesn't only share with (2.45) but also with all sentences on our list that have occurrences of the name *John*.[25] But the difference that matters is that (2.6.e) has names in the complement of attitudinal verbs.

What has been said so far about the interpretation of names in MSDRT on the one hand and about the interpretation of attitude attributions on the other strongly suggests, I take it, how names in attitudinal complements should be handled. And indeed, I think that for many cases this is just right. But it doesn't seem to be right for all cases, and questions arise in particular about examples of attributions with proper names that have been prominent in the philosophy of logic snd language.(Although these puzzles are not crucial for the applications we are going to make of MSDRT in our project, some of these puzzles are so well-known and still so often and hotly debated, that some discussion of them may be right, as a way of showing what benefits MSDRT can bring in those areas too.)

The suggestion for how to deal with names in attitude attributions from what we have so far said about MSDRT is this. In order that the speaker S of a sentence like (2.6.e) can make use of the names occurring in it she must have N-labeled entity representations for each name N occurring in her utterance. That is so whether N occurs inside an attitude complement, as *Mary* and *Paris* in (2.6.e), or not as for *John*. But the cases of interest are *Mary* and *Paris* and this is the core of the suggestion: By using these names in her attribution to John, S conveys that the belief she attributes

[25]If we take the treatment of proper names exemplified by that given above for *Gogol* as the proper way to deal with names in general, then all our earlier treatments of the sentences on our list are up for revision. I hope it is clear from the discussion around (2.45) how that can be done and also that this would not lead to different predictions about the truth conditions of those sentences.

to him is about the bearers of those names – that its content is a doubly singular proposition. And that, according to MSDRT means that John must have entity representations for those bearers. It does *not* entail, however, that John must know those bearers under the names S is using. In fact, it is possible that his entity representations are without any labels at all.

We argued above that to interpret utterances with standard uses of names the interpreter H must also have entity representations for the names the speaker uses (and if not, he must accommodate some).

So we may assume that the relevant part of H's mental state after interpretation of (2.6.e) is as in (2.53).

$$
(2.53)\left\{
\begin{array}{l}
\left\langle [ENT, j], \begin{array}{|c|}\hline \\ \hline \text{person}(j) \\ \text{Named}(j, John) \\ \hline \end{array}, \mathcal{K}_j \right\rangle \\[2em]
\left\langle [ENT, m], \begin{array}{|c|}\hline \\ \hline \text{person}(m) \\ \text{Named}(m, Mary) \\ \hline \end{array}, \mathcal{K}_m \right\rangle \\[2em]
\left\langle [ENT, p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p, Paris) \\ \hline \end{array}, \mathcal{K}_p \right\rangle \\[2em]
\left\langle BEL, \begin{array}{|c|}\hline s \quad t \\ \hline t = n \quad t \subseteq s \\
s\text{: } Att \left(j, \left\{ \begin{array}{l}
\left\langle [ENT, w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline \end{array}, \mathcal{K}_w \right\rangle \\[1em]
\left\langle [ENT, c], \begin{array}{|c|}\hline \\ \hline \text{city}(c) \\ \hline \end{array}, \mathcal{K}_c \right\rangle \\[1em]
\left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: loc}(w,c) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{l} < w, m >, \\ < c, p > \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

Note the links in the fourth argument position of *Att*. These connect John's ERs with the ERs that are constituents of H's own mental state MSD in (2.53). The effect of this is that the belief that (2.53) attributes to John is about the referents of H's own entity representations that enable him to refer to those entities as 'Mary' and 'Paris'. Intuitively this seems the right way for H to represent the content of the attributed belief, as the doubly singular proposition that the person represented by his own *Mary*-labeled entity representation is in the place represented by his own *Paris*-labeled entity representation. It will be useful, however, also in the light of considerations later on, to state explicitly what operations are needed as part of the interpretation of a name occurring in an attitude content specification that has been used by the source in what we have called the 'standard way' for such name uses. (2.54) spells this out.

(2.54) (Rule for the interpretation of the standard use of names in attitude contexts, also called the *Reference-Based Principle*)

Let γ be the complement clause or phrase of an attitude attributing verb V in an utterance that H is interpreting and assume that in the interpretation that H is building the combination $V + \gamma$ is represented as a Propositional Attitude $<$MOD,$K$$>$. More precisely, $<$MOD,$K$$>$ is a constituent of an MSD that occurs as 3rd argument of some *Att*-predication with an gent -representing dref a in 2nd argument position. Suppose further that γ contains an occurrence of a proper name N and that H takes the source S to have used N in the standard way. Then H has to perform the following operations as part of his interpretation of N:

(i) H either has to use an N-labeled entity representation he has and that he takes to represent the same entity as the N-labeled entity representation that S relied on in her use of N, or else accommodate such an entity representation. Let x_H be the distinguished dref of this entity representation.

(ii) H must make sure that the mentioned MSD contains an ER for the referent of N that is linked with his own entity representation via a link in the 4th argument position of the *Att*-predication; that is, the link set should contain the pair $<x_a,x_H>$, where x_a is the distinguished dref of this ER. If the MSD already contains such an ER with the described link, then nothing needs to be done in this step. If not, then H must accommodate an ER as a constituent of the MSD, and add the link to the link set in the 4th argument position of the *Att*-predication.

(iii) x_a is to be inserted in K into the argument position or positions that correspond to the argument position or positions of N in γ.

(end rule formulation)

It's time for a minor addition. In this last subsection we have been analyzing the semantic contents of utterances in terms of interpreters' beliefs. But why beliefs? Interpretation of what you hear or read involves two steps: first you need to identify what it is the speaker or text is saying. Only then, and on the basis of the content that you have thereby identified will you adopt what you are being told as a belief, or reject it false or attribute a greater or lesser probability to it. But the appeal that the communication-theoretic approach makes to the linguistic competence of the interpreters it considers has nothing to do with the question whether H accepts or rejects the content he has identified. It makes more sense to the bare result of interpretation, before it is subjected to doxastic assessment.

One way to make room for this in MSDRT is to adopt a new Mode Indicator *CONT*, which expresses this doxastic neutrality. We assume that *CONT* has an argument slot for the utterance that H has interpreted as having the given content. In other words, MSDs can now also have constituents of the form $<CONT(u),K>$, where K is H's interpretation of the utterance u, and it is assumed that u is the distinguished dref of H's ER for the utterance to which he has assigned K as interpretation.[26]

Important at this stage of interpretation is that the interpreter has a record of the utterance that the content identified by his 'attitude' is the representation of and that this is the utterance of which it is the representation. For $<$CONT,$K>$ to be a constituent of H's mental state signifies that H has identified this content but hasn't yet come to a decision as to what propositional attitude properly speaking to make out of it. Using this notation we can represent the first stage result of H's interpretation of a speaker S's utterance of (2.6.e) as in (2.55).

[26] Formal theories of meaning in language and linguistic communication have for the most part taken it for granted that utterance recipients decide, as part of their interpretations (a) what kind of speech act the utterance is and (b) what to do with that speech act; in particular, when they take the speech act to be an assertion, whether they accept its content as true, reject it as false, or remain undecided, waiting perhaps for further evidence going either way. But that is quite unrealistic. Many, perhaps even most of the utterances that reach us are interpreted by us to the point that we represent their content, but leave it at that because the utterance (with the content we have extracted from it) doesn't concern us. In those cases of this kind where we take the utterance to be an assertion, we don't move beyond a representation of the form $<CONT(u),K>$.

$$
(2.55)\left\{
\begin{array}{l}
\left\langle [ENT,u], \begin{array}{|c|}\hline \\ \hline \text{utterance}(u) \\ \text{Form}(u,\text{'John believes that Mary is in Paris'}) \\ \hline \end{array}, \mathcal{K}_u \right\rangle \\[2ex]
\left\langle [ENT,j], \begin{array}{|c|}\hline \\ \hline \text{person}(j) \\ \text{Named}(j,John) \\ \hline \end{array}, \mathcal{K}_j \right\rangle \\[2ex]
\left\langle [ENT,m], \begin{array}{|c|}\hline \\ \hline \text{person}(m) \\ \text{Named}(m,Mary) \\ \hline \end{array}, \mathcal{K}_m \right\rangle \\[2ex]
\left\langle [ENT,p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p,Paris) \\ \hline \end{array}, \mathcal{K}_p \right\rangle \\[2ex]
\left\langle CONT(u), \begin{array}{|c|}\hline s \quad t \\ \hline t = n \quad t \subseteq s \\
s\text{: } Att \left(j, \left\{ \begin{array}{l}
\left\langle [ENT,w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline \end{array}, \mathcal{K}_w \right\rangle \\
\left\langle [ENT,c], \begin{array}{|c|}\hline \\ \hline \text{city}(c) \\ \\ \hline \end{array}, \mathcal{K}_c \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: loc}(w,c) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{l} < w, m >, \\ < c, p > \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

The *CONT* constituent of (2.55) will then be the point of departure for various possible attitudes towards the content of this constituent.

2.2.5 Iterated Attributions

Att-predications can occur as part of MSDs that in their turn fill argument positions of *Att* predicates. This makes it possible to represent sentences that attribute to one agent thoughts about yet another agent, and so on down.

Here are some examples of such sentences.

(2.56)a. John believes that Mary thinks that Bill is sick.

b. John believes that Mary thinks that he is sick.

c. John believes that Mary thinks that she is sick.

d. John believed that Mary thought that Bill was sick.

e. John hopes that Mary thinks he is a hero. He fears that she thinks he is a nincompoop.

f. John thinks that Mary thinks that he thinks that she has eaten the last doughnut.

The Logical Form for 2.56.a) is shown in (2.57). (Once again the representation is split into two parts for space reasons, with the representation of Mary's mental state in (2.57.b) and its abbreviation 'BEL$_{Mary}$' holding its position in (2.57.a).)

The names are treated the same way as those of (2.6.e) in the last section, except that at the 'top level' we have returned to the original mode of representation, consisting of a dref x_N to represent the referent of the name N and the predication 'N'(x_N)'.

(2.57)a.

$$\begin{array}{|c|}\hline s \quad t \quad j \quad m \quad b \\ \hline t = n \quad t \subseteq s \quad \text{John'}(j) \quad \text{Mary'}(m) \quad \text{Bill'}(b) \\ s\text{: } \mathit{Att} \left(j, \left\{ \begin{array}{l} \left\langle [ENT, w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline \end{array}, \mathcal{K}_w \right\rangle \\ \\ \left\langle [ENT, v], \begin{array}{|c|}\hline \\ \hline \text{person}(v) \\ \\ \hline \end{array}, \mathcal{K}_v \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ \text{BEL}_{Mary} \\ \hline \end{array} \right\rangle \end{array} \right\}, \left\{ \begin{array}{l} < w, m >, \\ < v, b > \end{array} \right\} \right) \\ \hline \end{array}$$

b. $$\text{BEL}_{Mary} = s''\text{: } \mathit{Att} \left(w, \left\{ \begin{array}{l} \left\langle [ENT, v'], \begin{array}{|c|}\hline \\ \hline \text{person}(v') \\ \hline \end{array}, \mathcal{K}_{v',w} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline s^3 \\ \hline n \subseteq s^3 \\ s^3\text{: sick'}(v') \\ \hline \end{array} \right\rangle \end{array} \right\}, \{< v', v >\} \right)$$

The reading of (2.56.b) that I want to focus on is that in which *he* is anaphoric to *John* and the belief attributed to John is understood as a self-attribution with regard to the argument position held by *he*. In MSDRT this entails that according to John's belief Mary has an entity representation for him and that it is of the individual represented by that entity representation that she thinks that he is sick. The MSDRT representation of this reading of (2.56.b) looks very similar to that for (2.56.a). The only significant differences are the links.

(2.58)a.

$$\begin{array}{|c|}\hline s\quad t\quad j\quad m \\ \hline t = n \quad t \subseteq s \quad \text{John'}(j) \quad \text{Mary'}(m) \\ s\text{: } Att\left(j, \left\{ \begin{array}{l} \left\langle [ENT,w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline\end{array}, \mathcal{K}_w \right\rangle \\ \\ \left\langle [ENT,v], \begin{array}{|c|}\hline \\ \hline \text{person}(v) \\ \\ \hline \end{array}, \mathcal{K}_v\right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n\subseteq s' \\ \text{BEL}_{Mary} \\ \hline\end{array}\right\rangle \end{array}\right\}, \{< w,m >\}\right) \\ \hline \end{array}$$

b.

$$\text{BEL}_{Mary} = s''\text{: } Att\left(w, \left\{ \begin{array}{l} \left\langle [ENT,v'], \begin{array}{|c|}\hline \\ \hline \text{person}(v') \\ \hline\end{array}, \mathcal{K}_{v',w} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline s^3 \\ \hline n\subseteq s^3 \\ s^3\text{: sick'}(v') \\ \hline\end{array}\right\rangle \end{array}\right\}, \{< v',i >\}\right)$$

Crucial in this representation is the link $<v',i>$ in the 4th argument position of the *Att*-predication that describes Mary's mental state. It is this link that makes the description provided by this *Att*-predication one that ascribes to Mary a belief that John takes to be about himself.

Exercises 1. Give the DRS for (2.56.c). 2. Construct the DRSs for (2.56.a,b,) given in 2.57) and 2.58) and your own DRS for (2.56.c) on the basis of syntactic trees for the three sentences.

2.2.6 I, *you* and i

MSDRT, we have seen, uses the special dref for i to represent an agent's self when it occurs in an MSD used to describe a mental state of that agent. There is an obvious connection between this function of i and the first person pronoun I: When a speaker uses I this is always to express a content that she herself represents using a i. (Or, to put the matter in slightly different terms, when a speaker interprets her own words – for instance, after having put them down on paper and then reads them again; but also when she pays attention to the words she produces in a way that her interlocutors will – then she will represent her own productions of I as instances of i.

This doesn't mean that speakers never refer to themselves with other DPs than I. On the one hand people may use *yours truly* or use their own name the way Julius Caesar does in *De Bello Gallico*, to express thoughts that they themselves represent with the help of i. And on the other hand there are the cases where one refers to oneself thinking one is referring to someone else, for instance when you comment, non-jokingly, on a recording of someone singing in your mother tongue by saying: "This man might be fine, if only he would stick to his own language." Here the speaker expresses a thought about someone for whom he has an anchored entity representation, with a perceptual anchor testifying to the auditory experience he has just had: We can have representations of ourselves of which we mistakenly assume that they represent somebody else. In such cases the speaker *couldn't* have used I as referring expression.

But such examples are special, confirming that when we want to refer to ourselves, from our own internal perspective, then the expression to use is the first person pronoun I.

The flip side to this is the response to I by the recipients of utterances containing it. Plainly, you cannot represent the individual the speaker is referring to by i in the way she uses it, viz. as a way to make self-attributions.[27] What we do, rather, when we interpret the uses of I by those addressing us is to attribute the properties she self-ascribes to (the referent of) the entity representation we have of her, as the one who is speaking to us.

Mental representations involving i also stand in a close systematic relation to the second person pronoun *you*, but one that is the reverse, you might

[27]Apparently this is a mistake that is sometimes made by young children; but they have to be quite young and they won't make the mistake for very long.

say, of the relation between i and I. When a speaker uses *you* in addressing her addressee, then she does so to refer to a constituent of her thought that takes the form of an entity representation, with a perceptual anchor reflecting her present awareness of its referent as the one she is speaking to. But at the same time her using *you* is an invitation to the addressee to use i in representing her uses of *you*. In our terms: the representation he constructs of her utterance should have i where her own representation has occurrences of the distinguished dref of her entity representation for him.

The upshot of this is that when communication involving either I or *you* is optimally successful – those occasions where interpretation leads to a representation that matches the speaker's representation in its truth conditions (and also as much as possible in other respects) – the two representations will have to be different: the interpreter's representation must have the distinguished dref of an entity representation for the speaker where the other has i.[28]

The give just one simple example of what this comes to in MSDRT, consider (2.59).

(2.59) I like you.

Assume that (2.59) is said by S to H and that S is speaking sincerely. Then S's mental state directly before and directly after her utterance will be describable as in (2.60.a). (2.60.b) gives the relevant representation of H's mental state after interpretation of S's words.

(2.60)a.
$$\left\{ \begin{array}{l} \left\langle [ENT, h_S] \ , \begin{array}{|c|}\hline \\ \hline \text{person}(h_S) \\ \hline \end{array} \ , \ \mathcal{K}_{h_S} \cup \begin{array}{|c|}\hline s^1 \\ \hline n \subseteq s^1 \\ s^1\text{: see}(i,h_S) \\ \hline \end{array} \right\rangle \\ \\ \left\langle BEL \ , \begin{array}{|c|}\hline s^2 \\ \hline n \subseteq s^2 \\ s^2\text{: like'}(i,h_S) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

[28]For detailed discussion see (Maier 2016). The basic observation that I and *you* differ in the way indicated is pretty obvious and trying to track it back to its earliest documentation in writing might well be a wild goose chase. I mention Maier's paper because it deals with the issue in a sprit akin to the communication-theoretic one adopted in these notes.

b.
$$\left\{ \begin{array}{l} \left\langle [ENT, sp_H] , \begin{array}{|c|} \hline \\ \hline \text{person}(sp_H) \\ \hline \end{array} , \mathcal{K}_{sp_H} \cup \begin{array}{|c|} \hline s^3 \\ \hline n \subseteq s^3 \\ s^3\text{: see}(i,sp_H) \\ \hline \end{array} \right\rangle \\ \\ \left\langle BEL , \begin{array}{|c|} \hline s^4 \\ \hline n \subseteq s^4 \\ s^4\text{: like'}(sp_H,i) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

Note the by now familiar heuristics: the subscripts of the distinguished drefs indicate the agent whose mental state contains the ER of which they are the distinguished drefs. The lower case letters ‘s’ and ‘a’ indicate that the ERs in question stand for the speaker S and the addressee H, respectively.

The crucial difference between the MSDs in (2.60.a) and (2.60.b) is that between the Condition ‘s^2: like’(i,a_S)’ of (2.60.a) and ‘ s^4: like’(s_H,i) of (2.60.b). In the first the self is the subject of *like*, in the second it is the direct object.

Among the DRS Construction rules needed for the construction of the belief in (2.60.b) are in particular the one for occurrences of *I* and the one for occurrences of *you*. It is intuitively clear what these rules come to and I have already described them informally. To repeat – without going into formal details, which are left to a suitable documentation enterprise, in the spirit of (Kamp 2015), but for MSDRT rather than just for DRT – (i) the Construction Rule for *you* is simply to represent its occurrences in argument positions by i. (ii) The Construction Rule for *I* is more involved. It requires the interpreter to instantiate occurrences of *I* in S’s utterance by the distinguished dref of the ER that he will have of the speaker, as the one who has produced the utterance he is interpreting. In (2.60.b) we assumed the this was the ER that’s H of S, as the person that is sitting across from him. This won’t cover all possible communication situations, but we’ll leave it at this.[29,30]

[29] It could be argued that whenever an interpreter is interpreting an utterance, that entails that he has an entity representation for this utterance and with it one for its source. And that it is always the distinguished dref of his entity representation for the source that he has to inserted in the argument positions filled by *I* in the utterance. The cases where he can identify the source as a person he can see and see to be the one who produces the utterance would then constitute a subclass of this larger class.

[30] The handling of *I* and *you* by the speaker in utterance production must involve pro-

Along the lines shown in (2.60.a) and (2.60.b) for (2.59), MSDRT can also deal with more complex sentences involving *I* and *you*. A first example is (2.61).

(2.61) You think that I do not like you.

(**??**) gives the MSD we assume for the speaker S immediately before and immediately after her utterance of (2.61). (2.61.a) gives most of this MSD, (2.61.b) the part abbreviated in (2.61.a) as 'BEL$_H$'.

(2.62)

a.

$$\left\{ \begin{array}{l} \left\langle [ENT, h_S]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(h_S) \\ \hline\end{array}\ , \mathcal{K}_{h_S} \cup \left\{ \begin{array}{|c|}\hline s^1 \\ \hline n \subseteq s^1 \\ s^1\text{: see}(i,h_S) \\ \hline \end{array} \right\} \right\rangle \\ \left\langle BEL\ , \begin{array}{|c|}\hline s^2 \\ \hline n \subseteq s^2 \\ s^2\text{: } Att\left(h_S, \left\{ \begin{array}{l} \left\langle [ENT, sp_h], \begin{array}{|c|}\hline \\ \hline \text{p'n}(sp_h) \\ \hline \end{array}, \mathcal{K}_{sp_h} \right\rangle \\ \langle BEL, BEL_h \rangle \end{array} \right\}, \{ < sp_h, i > \} \right) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

duction rules that are mirror images of sorts of the interpretation rules just mentioned. Since there is at present no good account of utterance production in MSDRT, we have no way of saying more about these rules directly. However, here too it is possible to appeal to the principle that producers are among the interpreters of their own productions. But when it comes to *I* and *you*, their interpretation rules must of course be different from the rules for the addressee; in fact they are their exact reverse. (And by the same token, for a bystander, who is neither the speaker nor the addressee the rules will be different again. But all this is blatantly obvious and everyone will be able to work out further details if he is thus disposed.)

b. BEL$_h$:

$$\begin{array}{|c|}\hline s^4 \\ \hline n \subseteq s^4 \\ s^4\text{: } \neg \begin{array}{|c|}\hline s^5 \\ \hline s^5 \subseteq s^4 \\ s^5\text{: like'}(sp_h,i) \\ \hline\end{array} \\ \hline\end{array}$$

Legenda

1. The most important about the two representations above is the heuristics we have adopted in choosing the distinguished drefs for the different ERs. The symbols we have chosen are meant to make these representations easier to read but without explanation the effect may well be the opposite. The general strategy we have used is to use as main symbol one that can be understood as an acronym for the referent of the ER and the subscript as an acronym for the agent whose ER it is. Thus h_S is the distinguished dref of the ER that the speaker S has for the hearer H, and sp_h of the ER that the hearer H has for the speaker S. (We use '*sp*' here rather than '*s*' in order to avoid confusion with drefs for states.

2. Running out of symbols of the kind one would like to use happens all tto often and all too soom. We want to use the symbol '*s*' for all drefs that represent states. That requires some kind of indexation as soon as you need more than one state dref. Primes will do fine, but only for as long as there are no more than three. When there are more, strings of them become a blur, and if you really want to know which is which, you have put your nose close to the sheet or the screen and start counting; and that is not conducive to quick understanding. Numerical affixes do not have this problem. But subscripts on drefs are used for other purposes, so here we have resorted, for the first time in this text, to superscripts. These serve no other purpose than to keep the drefs they superscribe distinct.

3. The expression of negation in (2.62.b) is rather round-about. The DRS containing the negation sign expresses that the hearer doesn't like the speaker at the relevant time – here this is the utterance time n – it says that there is a state s^4 surrounding n which contains no substate s^5 to the effect that the hearer likes the speaker. This is the official form for expressing negation in DRT. It is motivated in part the aspectual properties of negation. (Negated clauses typically have imperfective aspect, see (de Swart & Molendijk 1999), (Kamp et al. 2011).) [end Legenda]

Once again the Logical Form that H will have constructed from (2.59) if he has interpreted the utterance the way it is meant to be is like (2.59), except for the switches between i and the distinguished drefs of the ER that each has for the other. (Compare (2.60a) and (2.60b).)

The step that leads from the representation construction for (2.59) to that for (2.61) can be iterated arbitrarily, leading to semantic representations for ever more deeply nested attributions. Just to give an impression of MSDRT's ability to handle multiply nested attributions and of the expressive power of its representation languages that is needed for that, we look at one more example, which involves embedding depth 2. (This is yet another sentence from the near inexhaustible repertoire of the tormented heart.)

(2.63) You think that I think that you do not like me.

Let us once more focus on the relevant part of the mental state of S, like we did in our discussion of (2.61). The 'outer shell' of the Logical Form we gave for (**??**) is the same as for (2.61). But the representation abbreviated as 'BEL$_h$' in (2.61.a) is now of course more complex. Instead of a belief content to the effect that the one isn't liked by the other, the content is now itself an attitude attribution to the other, of a belief the target of which is oneself. We refer to the new insert into (2.61.a), which captures this more complex belief on the part of S, as 'BEL'$_h$'. Here it is.

(2.64)

$$
\begin{array}{|c|}
\hline
s' \\
\hline
n \subseteq s' \\
s'\colon Att \left(sp_h, \left\{ \begin{array}{l} \left\langle [ENT, h_{sp(h)}], \begin{array}{|c|}\hline \\ \hline \text{p'n}(h_{sp(h)}) \\ \hline \end{array}, \mathcal{K}_{h_{sp(h)}} \right\rangle \\ \left\langle BEL, \begin{array}{|c|}\hline s^4 \\ \hline n \subseteq s^4 \\ s^4\colon \neg \begin{array}{|c|}\hline s^5 \\ \hline s^5 \subseteq s^4 \\ s^5\colon \text{like'}(h_{sp(h)}, sp_h) \\ \hline \end{array} \\ \hline \end{array} \right\rangle \end{array} \right\}, \left\{ < h_{sp(h)}, h_S > \right\} \right) \\
\hline
\end{array}
$$

Once again an explanation is called for in connection with the symbol we have chosen for the distinguished dref of the ER in this representation. The thought behind this choice, of the dref symbol '$h_{sp(h)}$', is that the ER in question is an Entity Representation that H has for S *according to the way she has represented his state of mind*. This is what $h_{sp(h)}$ is meant to convey. The dref represents H as represented by S according to what H thinks according to S.

This idea behind the choice of $h_{sp(h)}$ also accounts for the link $< h_{sp(h)}, h_S >$ in the 4th argument position of the *Att*-Condition in (2.64). Recall that the outer shell, as we just called it, of the Logical Form for (**??**) that was presented in (2.62.a) was a representation of the relevant part of the mental state of S. if we reuse this shell here, by inserting (2.64) for 'BEL_h' in (2.62.a), then we have once again a representation of the relevant part of the mind of S. Whose mental state is described by an MSD imposes certain constraints on the form that its links can take. For the link in (**??**) the justification runs like this. The content of the belief that S attributes to H in (**??** is an attribution that he makes to her of a belief that she attributes to him. (2.64) is the part of S's MSD that represents H's representation of her attribution to him. One constituent of that representation is the Entity Representation in the MSD that H attributes to S, with its distinguished dref $h_{sp(h)}$. Since this ER is what he takes to be her Entity Representation for him, its distinguished dref must be linked to him in some way. But what should that come to in this case? (2.62.a), which we are reusing in the MSD for the relevant part of S's mental state in relation to (2.63), has S's ER for H (with its distinguished dref h_S). There is only one way in which S can account for the intuition that according to what she thinks (and (2.63) expresses) H assumes that she has a belief about him (viz. that he doesn't like her). And that is to link the dref that occurs in the content representation of this belief she thinks he attributes to her – in other words, the dref $h_{sp(h)}$ that occurs as argument in the predication 's^5: like'$(h_{sp(h)}, sp_h)$' – with the agent h_S of the *Att*-predication that contains his representation of her attribution to him.

By inserting the link $< h_{sp(h)}, h_S >$ into the link set slot of the *Att*-Condition in (2.64) S expresses that from her point of view the belief that H attributes to her is a self-ascription on his part. She cannot do better than this because she is not him.

To see this point more clearly, let us also have a brief look at the mental state after having computed a logical Form for (2.63) and accepted it as a belief. The 'outer shell'will now be as in (2.65)

(2.65)

$$\left\{ \begin{array}{l} \left\langle [ENT, sp_H]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(sp_H) \\ \hline \end{array}\ , \ \mathcal{K}_{sp_H} \cup \left\{ \begin{array}{|c|}\hline s^1 \\ \hline n \subseteq s^1 \\ s^1\text{: see}(i, sp_H) \\ \hline \end{array} \right\} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|} \hline s^2 \\ \hline n \subseteq s^2 \\ s^2\text{: } Att \left(sp_H, \left\{ \begin{array}{l} \left\langle [ENT, h_{sp(h)}], \begin{array}{|c|}\hline \\ \hline \text{p'n}(h_{sp(h)}) \\ \hline \end{array}, \mathcal{K}_{h_{sp(h)}} \right\rangle \\ \\ \langle BEL, BEL_{sp} \rangle \end{array} \right\}, \left\{ < h_{sp(h)}, i > \right\} \right) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

Note that the missing part BEL_{sp} now has one level of embedding depth more than the representation in (2.64).It will have to be a representation that contains the MSD in 2.62.a) with (2.64) substituted for BEL_h. But there will be one difference to this representation. The ER for H in 2.62.a) will now be what H takes to be S's representation for him. That is, the distinguished dref h_S of this ER will no be linked to H's self. This link then forms a (very short) chain with the link $<h_{sp(h)}, h_S>$, thereby linking $h_{sp(h)}$ to H's i. So for him the attribution that S makes to him is a genuine self-attribution: What S accuses him of by saying (2.63) to him is that thinks she has a certaim thought – the thought that he doesn't like her – and that that thought is directly aout him is directly about him.

One upshot of all this is that utterances which contain both *I* and *you* will be understood by both speaker and addressee as self-ascriptions, but withv respect to different roles in the predication structures that are exoressed in the sentence.

A further challenge by sentences like those we have looked at in this section is to formulate exact interpretation rules for *I* and *you* and to see them in action in the interpretation of sentences like these. We. leave this task for later.

2.3 Abbreviations

MSDRT's Logical Forms for attitude reports are motivated largely by the problem of multiple attitude attributions. When can a presupposition generated by the content specification of one attribution be resolved on the basis of the content of some other attribution? MSDRT deals with this problem by treating every attribution as a partial description of the attributee's mental state (at some time) and then providing mechanisms for deciding which attributions are descriptions of the same mental state; and if so, for determining which attitudes belonging to that state can serve as contexts for the resolution of presuppositions triggered by which others.

This may be the right way to deal with this aspect of multiple attributions. And it can never be wrong to treat an attitude attribution as part of some larger mental state. But to write this out, for each and every attribution, to the last title and iota the way we have so far done soon feels like an awkward burden and one that often impedes a better understanding rather than assisting it. It is therefore going to be helpful to introduce various abbreviations. These will also make it easier to compare MSDRT with other approaches to the theory of attitudes and the semantics of those attitude reports for which those approaches make some clear predictions.

The first abbreviation proposal is for a very simple type of attitude attribution, whose content representation is DRS without free drefs. That is, the full logical form of the attribution is as in (2.66).

$$\text{(2.66)}\quad \begin{array}{|c|} \hline s \\ \hline t \subseteq s \\ s\text{: } \mathit{Att}\left(x, \left\{ \langle MOD, K \rangle \right\}, \emptyset\right) \\ \hline \end{array}$$

This is the abbreviation we will use for Logical Forms like (**??**):

$$\text{(2.67)}\quad \begin{array}{|c|} \hline \\ \hline \langle MOD_{x,t}, K \rangle \\ \hline \end{array}$$

(2.66) being an abbreviation of (2.66) is to be understood as follows: Suppose a DRS K^1 has a DRS K^2 of the form (2.66) as a part, in the sense that there is a DRS K^3 such that $K^1 = K^2 \bigoplus K^3$. Then K^1 can be simplified to the result K^4 of merging K^3 with the instance K^5 of (2.67) abbreviating K^2:

$$K^4 = K^3 \bigoplus K^5$$

For an example, consider the Logical Form (2.12.c), repeated here

$$\text{(2.12.c)} \quad \begin{array}{|c|}\hline t \quad s \quad j \\ \hline t = n \quad t \subseteq s \quad \text{John'}(j) \\ s\colon Att\left(j, \left\{\left\langle BEL, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\colon \text{PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\colon \text{rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \right\}, \emptyset \right) \\ \hline \end{array}$$

Abbreviating (2.12.c) in the manner of (2.67) leads to (2.68).

$$\text{(2.68)} \quad \begin{array}{|c|}\hline t \quad j \\ \hline t = n \quad \text{John'}(j) \\ \left\langle BEL_{j,t}, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\colon \text{PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\colon \text{rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \\ \hline \end{array}$$

This abbreviation option can be extended further. For example, to Propositional Attitude constituents $\langle MOD, K \ \rangle$ of DRSs in which K contains free drefs.

The most important case of this for practical purposes is of drefs α that are the distinguished dref of some ER in the same MSD as $\langle MOD, K \ \rangle$, where this MSD fills the third argument position of some *Att*-predication and where

α also occurs as first member of some link $<\alpha,\beta>$ in the fourth argument position of *Att*. Suppose that the MSD consists just of $\langle MOD, K \rangle$ and ERs whose distinguished drefs $\alpha_1,\ldots,\alpha_n$ occur in K. Then the *Att*-predication may be abbreviated as illustrated by the following example, going back to our discussion of 'John believes that Mary is in Paris.' Recall the representation in (2.53) of the addition to H's mental state as a result of his interpretation of this sentence.

$$(2.53)\quad \left\{ \begin{array}{l}
\left\langle [ENT, j], \begin{array}{|c|}\hline \\ \hline \text{person}(j) \\ \text{Named}(j,\mathit{John}) \\ \hline \end{array}, \mathcal{K}_j \right\rangle \\[2ex]
\left\langle [ENT, m], \begin{array}{|c|}\hline \\ \hline \text{person}(m) \\ \text{Named}(m,\mathit{Mary}) \\ \hline \end{array}, \mathcal{K}_m \right\rangle \\[2ex]
\left\langle [ENT, p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p,\mathit{Paris}) \\ \hline \end{array}, \mathcal{K}_p \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline s \quad t \\ \hline t = n \quad t \subseteq s \\
s\text{: } \mathit{Att} \left(j, \left\{ \begin{array}{l}
\left\langle [ENT, w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline \end{array}, \mathcal{K}_w \right\rangle \\
\left\langle [ENT, c], \begin{array}{|c|}\hline \\ \hline \text{city}(c) \\ \\ \hline \end{array}, \mathcal{K}_c \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: loc}(w,c) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{l} <w, m>, \\ <c, p> \end{array} \right\} \right) \\ \hline \end{array} \right\rangle
\end{array} \right\}$$

Before we proceed with spelling out the abbreviation there is a matter that could and arguably should have been addressed earlier. The ERs in the MSD of the *Att*-predication in (2.53) were introduced on the basis of the principle that the use of the proper names *Mary* and *Paris* (in the content specification of the belief attribution to John that S has made and H has interpreted) entails that John must have had, in conjunction with the belief attributed to him, entity representations for the bearers of these names. But the precise form of the ERs describing these entity representations was never properly justified so far. We just assumed that John would have representations of these two entities that assign them to the right ontological categories ('person' in the first and 'city' in the second case); and for the anchor sets we used the schematic symbols $\mathcal{K}_w$ for the first and $\mathcal{K}_c$ for the second representation. The implicit assumption behind these symbols was that they convey nothing about what these anchor sets in John's mind really were. In the form in which it is made, the report offers no information about these components of John's entity representations for *Mary* and *Paris*. From the perspective of truth conditions this is right: Form that perspective what matters about the ERs in the MSD of John's mental state is just that of making sure that his belief is about the entities represented by the drefs m and p in the anchors of the *Att*-Condition.

In light of these considerations it should be clear that no essential loss of information results from simplifying (2.53) to (2.69).

$$
(2.69)\left\{
\begin{array}{c}
\left\langle [ENT, j], \begin{array}{|c|}\hline \\ \hline \text{person}(j) \\ \text{Named}(j, John) \\ \hline \end{array}, \mathcal{K}_j \right\rangle
\left\langle [ENT, m], \begin{array}{|c|}\hline \\ \hline \text{person}(m) \\ \text{Named}(m, Mary) \\ \hline \end{array}, \mathcal{K}_m \right\rangle \\[2ex]
\left\langle [ENT, p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p, Paris) \\ \hline \end{array}, \mathcal{K}_p \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline t \\ \hline t = n \\ \left\langle BEL_{j,t}, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: loc}(m,p) \\ \hline \end{array} \right\rangle \\ \hline \end{array} \right\rangle
\end{array}
\right\}
$$

The last constituent of H's MSD is the abbreviated representation of H's belief that John currently believes that Mary is currently in Paris. The belief attributed to John is explicitly represented as about the entities for which H has representations by the *Mary*- and *Paris*-labeled ERs in (2.69). The MSDRT principle that this is only possible when John has entity representations for these entities himself is not abandoned by the way in which (2.69) abbreviates (2.53) is not given up, but it is now only implicit.

We can extend the use of abbreviations also in another direction. Recall (2.6.c), the attribution to John of a belief and a hope with a content that is dependent on the belief content for the resolution of its *stop*-presupposition. We already saw how the abbreviation of the belief conjunct of this report in (2.68), repeated here.

(2.68)
$$\begin{array}{|c|}\hline t \quad j \\ \hline t = n \quad \text{John'}(j) \\ \left\langle BEL_{j,t}, \begin{array}{|c|}\hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: PROG}(\hat{}e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \\ \hline \end{array}$$

We can abbreviate the preliminary representation of the hope conjunct in the same way. This simplifies this representation – shown earlier under (2.27) and also repeated here – to (2.70). Note that, given the way in which the preliminary representation for the second conjunct was displayed in (2.27.a,b), the only change is in the part that contains the Mode Indicator $HOPE$.

(2.27)

a. $$\left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline z? \\ \hline \text{male}(z) \\ \hline \end{array}, \\ \\ \begin{array}{|c|}\hline r?\ \rho? \\ \hline \rho(r,t'') \\ \hline \end{array} [pres] \end{array} \right\}, \begin{array}{|c|}\hline t'' \quad s'' \\ \hline t'' = n \quad t'' \subseteq s'' \\ s''\text{: } Att\Big(z, \Big\{ \ \langle HOPE, [DP] \rangle \ \Big\}, \emptyset\Big) \\ \hline \end{array} \right\rangle$$

$$\text{b. DP} = \left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline s''? \\ \hline s''\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array}, \\ \\ \begin{array}{|c|}\hline r'?\ \rho? \\ \hline \rho'(r',t''') \\ \hline \end{array} [fut] \end{array} \right\}, \begin{array}{|c|}\hline t'''\quad e\quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'',s^4) \\ (s'' \supset\subset e \supset\subset s^4) \\ s^4\text{: } \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s^4 \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle$$

(2.70)

$$\text{a. } \left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline z? \\ \hline \text{male}(z) \\ \hline \end{array}, \\ \\ \begin{array}{|c|}\hline r?\ \rho? \\ \hline \rho(r,t'') \\ \hline \end{array} [pres] \end{array} \right\}, \begin{array}{|c|}\hline t'' \\ \hline t'' = n \\ \langle HOPE_{z,t''}, [DP] \rangle \\ \hline \end{array} \right\rangle$$

$$\text{b. DP} = \left\langle \left\{ \begin{array}{l} \begin{array}{|c|}\hline s''? \\ \hline s''\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array}, \\ \\ \begin{array}{|c|}\hline r'?\ \rho? \\ \hline \rho'(r',t''') \\ \hline \end{array} [fut] \end{array} \right\}, \begin{array}{|c|}\hline t'''\quad e\quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'',s^4) \\ (s'' \supset\subset e \supset\subset s^4) \\ s^4\text{: } \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s^4 \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle$$

The question whether the belief content can be used in justifying the *stop*-presupoposition can now be asked and answered in the same way as before: Is it possible to show that the $HOPE$ constituent in (2.70) and the BEL constituent in (2.68) are constituents of one and the same mental state? And the answer: Yes, if we can show that the subscripts of the two Mode Indicators, $HOPE_{z,t''}$ and $BEL_{j,t}$ coincide. And that has already been done: t'' and t can both be shown to coincide with n and z can be (and has to be) resolved to j.

$$`\left\langle BEL_{j,t}, \begin{array}{|c|} \hline t' \quad s' \\ \hline t' = n \quad t' \subseteq s' \\ \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: rain'} \\ \hline \end{array}\ ) \\ \hline \end{array} \right\rangle'$$

and

$$`\left\langle HOPE_{z,t''}, \begin{array}{|c|} \hline t''' \quad e \quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'', s^4) \\ (s'' \supset\subset e \supset\subset s^4) \\ \\ s^4\text{:} \begin{array}{|c|} \hline \\ \hline \neg \begin{array}{|c|} \hline s^3 \\ \hline s^3 \subseteq s^4 \\ \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|} \hline e \\ \hline e\text{: rain'} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle'$$

Once that has been done and the presuppositions of (2.70) have been resolved as before, the two attitudes attributed to John in (2.6.c) can be represented as the DRS Conditions below that will cooccur in the Condition Set of an abbreviated Logical Form for this attribution. The resulting representation is shown in (2.71).

(2.71)
$$\begin{array}{|c|}\hline j \quad z \quad t \quad t'' \\ \hline \text{John'}(j) \quad t = n \quad t'' = n \quad z = j \\ \left\langle BEL_{j,t}, \begin{array}{|c|}\hline t' \;\; s' \\ \hline t' = n \quad t' \subseteq s' \\ s'\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array}) \\ \hline \end{array} \right\rangle \\ \left\langle HOPE_{z,t''}, \begin{array}{|c|}\hline t''' \quad e \quad s^4 \\ \hline n_e \prec t''' \quad e \subseteq t''' \\ e\text{: Trans}(s'',s^4) \\ (s'' \supset\subset e \supset\subset s^4) \\ s^4\text{:} \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline s^3 \\ \hline s^3 \subseteq s^4 \\ s^3\text{: PROG}(^\wedge e. \begin{array}{|c|}\hline e \\ \hline e\text{: rain'} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \hline \end{array} \right\rangle \\ \hline \end{array}$$

In view of the equalities '$t = n$', $t'' = n$' and '$z = j$' the Belief Condition and the Hope Condition belong to the same mental state (same agent, same time). So it is possible to reconstruct from (2.71) an MSD for the mental state of John's of which they are both constituents.

Abbreviations of this kind will become useful when we deal with the often iterated attitude descriptions that arise in joint planning. But sometimes it will be useful or necessary to return to the official MSDRT notation, in which *Att*-predications play their pivotal part. And it is always good to keep the official notation in mind.

The abbreviations introduced in this section have so far been limited to Propositional Attitudes. But there is no need for this limitation. ERs can be indexed in the same way, so that they too wear the mental states to which they belong on their sleeves. Take for instance the ERs in the *Att*-predication of (2.53), which is displayed here once more:

$$
(2.53)\quad \left\{
\begin{array}{l}
\left\langle [ENT, j], \begin{array}{|c|}\hline \\ \hline \text{person}(j) \\ \text{Named}(j, \mathit{John}) \\ \hline \end{array}, \mathcal{K}_j \right\rangle \\[2ex]
\left\langle [ENT, m], \begin{array}{|c|}\hline \\ \hline \text{person}(m) \\ \text{Named}(m, \mathit{Mary}) \\ \hline \end{array}, \mathcal{K}_m \right\rangle \\[2ex]
\left\langle [ENT, p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p, \mathit{Paris}) \\ \hline \end{array}, \mathcal{K}_p \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline s \quad t \\ \hline t = n \quad t \subseteq s \\
s\text{: } \mathit{Att}\left(j, \left\{ \begin{array}{l}
\left\langle [ENT, w], \begin{array}{|c|}\hline \\ \hline \text{person}(w) \\ \hline \end{array}, \mathcal{K}_w \right\rangle \\[1ex]
\left\langle [ENT, c], \begin{array}{|c|}\hline \\ \hline \text{city}(c) \\ \\ \hline \end{array}, \mathcal{K}_c \right\rangle \\[1ex]
\left\langle BEL, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: loc}(w,c) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{l} < w, m >, \\ < c, p > \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

We can indicate that they belong to the mental state of John at t by attaching the relevant drefs as subscripts to their first components. For instance, the first ER then becomes (2.72).

(2.72) $$\left\langle [ENT, w]_{j,t}, \begin{array}{|c|} \hline \\ \hline \text{person}(w) \\ \text{Named}(w, John) \\ \hline \end{array}, \mathcal{K}_w \right\rangle$$

Note that by extending this indexing device also to ERs we have provided ourselves with an alternative way of describing mental states: instead of using MSDs filling the third argument position of *Att* we index ERs and Propositional Attitudes with the drefs for the agent and the location time of the mental state. Sharing their indexations then shows two such DRS Conditions to be part of the same mental state.

A further point that will become centrally important: In the above we have concentrated on the case of two constituents $<MOD^1_{x,t},K^1>$ and $<MOD^2_{y,t'},K^2>$ that belong to the same momentary mental state. Is it the case, the question was, whether $x = y$ and $t = t'$? But it is often also important to be able to tell from our representations whether $<MOD^1_{x,t},K^1>$ and $<MOD^2_{y,t'},K^2>$ are part of the momentary mental states of the same agent, though at different times t and t'.

This last question touches on an aspect of the mental life that is as obvious as it is crucial: mental state constituents have durations; they exist in the minds of their agents for often extensive periods of time. In our terms this means that what can be described on the one hand as a constituent $<MOD^1_{x,t},K^1>$ of x's mental state at t and on the other hand as a constituent $<MOD^2_{x,t'},K^2>$ of x's mental state at t' is one and the same propositional attitude, whose existence includes both t and t'. In Ch. 4, Section 4.1 we will introduce the machinery to make it explicit that components $<MOD^1_{x,t},K^1>$ and $<MOD^2_{x,t'}$ of the mental states of an agent x at different times t and t' are 'slices' of the same Propositional Attitude (or Entity representation.[31]

The 'orthogonal' case – where x and y are different but t and t' are the same is important too., These are the cases where $<MOD^1_{x,t},K^1>$ and $<MOD^2_{y,t'},K^2>$

[31] It might be thought that if $<MOD^1_{x,t},K^1>$ and $<MOD^2_{x,t'},K^2>$ are time slices of the same mental state constituent, then MOD^2 should be the same as MOD^1. But it ia arguable that even in this regard the same attitude may change over time. This too is a point to be discussed at length later on. **[Do we?]**

are attitudes held by different agents x and y at the same time $t = t'$ – will often be of importance too, especially in situations of joint deliberation and planning, where it is essential for the deliberators that they arrive at shared plans and intentions. This is a point we will take up in Ch. 6. [**This hasn't yet been done.**]

Chapter 3

More on Proper Names in the Complement Clauses of Attitudinal Verbs

In the last chapter we adopted vicarious anchors to account for the assumption that we can acquire the ability to refer on the basis of the references we can observe are made by others. This covers among other things the transfer of names to new users according to the Causal Theory: by hearing you use a name N to refer to something, I can take over N from you as the name for whatever it was that you just used it to refer to. In MSDRT terms: My reaction to your use of N to refer to **d** is to introduce an N-labeled ER for **d** with a vicarious anchor that witnesses the causal connection between me and **d** that is the composition of the causal relation between me and your use of N and the presumed causal relation in which you stand to **d**, and which – a further assumption of MSDRT – takes the form of an N-labeled ER for **d** that you have.

According to this story vicarious anchors are witnesses of causal relations that connect agents with entities about which they can think and speak. But since the causal relations they witness involve the causal effects of communications between users, vicarious anchors are inevitably also witnesses to those communicational interactions that give rise to the ERs that they come along with. This communicative dimension to them is essential to MSDRT's reconstruction of the Causal Theory of Names: vicarious anchors may fail to provide their ERs with proper referents, because what the possessor of the ER took to be a proper reference by his interlocutor's use of N wasn't one. But even in such cases the vicarious anchor doesn't fail as witness of the communicative interaction as such, and the commitment to use N to refer to whatever it was that the interlocutor used it to refer to: Vicarious anchors are the creators of the links of causal chains (Kamp 2022*a*). This section is about the communication-theoretic dimension of vicarious anchors in various situations in which they do not function as true witnesses to real reference.

There are a number of different cases we want to consider. The first group consists of those cases in which the interpreter erroneously adopts a vicariously anchored ER in response to what he takes to be a reference made in the utterance he interprets. This group can be subdivided into the cases in which the speaker knows that she isn't properly referring by using a name or other referring expression and the cases where she doesn't. When she doesn't, then this will be because she herself was under an illusion that led her to adopt an ER with a defective vicarious anchor, or – and this will typically be the case when the expression she used is a proper name – because she herself is under the false impression that the name was used to properly refer when she got it from the source that she got it from. In this latter case it should in principle be possible to trace the use of the name back to one or more speakers who used it knowing that it didn't properly refer, which leads us back to an instance of the first case.

A speaker may knowingly use a name (or other expression designed for referential use) non-referentially for various reasons. She may invent a wealthy uncle in the Americas to obtain the credit that otherwise she would not have been granted, and get away with it. Or she may have used the name non-referentially in the way she knows others do and may think everyone does, like most of us will probably now use *Nessie* (the Loch Ness monster), or *The Abominable Snowman*; in these latter cases she may not think that her audience will mistake her use of the name as a case of proper reference; or she may not think about this possibility at all, or simply not care.[1]

And then there are the cases where the speaker doesn't refer by using a name, and is aware that her use of it will be taken to be referential and does nothing to prevent this in order to spare the addressee's feelings or illusions, as when mummy talks with Johnny about Santa: she knows there is no Santa, he doesn't and for the time being she knowingly lets this state persist. I will have more to say about this case further down.

[1]The use of fictional names in communication about the fictional worlds in which they name, and where all involved in the communication are fully aware that the names are names of people and things in the given fictional world, is a topic in its own tight. See for instance the relevant essays in (Maier & Stokke 2020).

In some such cases the interpreter will misconstrue the speaker's use of the name or other expression designed for reference as a case of genuine reference. But for the second group of cases to be considered now this isn't so. In these cases the interpreter is aware that the speaker used a name of other definite noun phrase without referring to some real entity.

Such communicative events also establish links between the speakers involved. For instance, the speaker may use a name that is unfamiliar to the recipient, the recipient may take the speaker to use the name without referring to a real referent and yet he may take over the name from her and may use it himself in this same non-referential way. This is how many of us learn names of fictional entities from others, even before they encounter them in the pieces of fiction where they originated. Here too there is a commitment on the part of the interpreter to use the name in the same way as the one from whom one gets it. But it is a commitment to 'coreference' of a different kind than when there is an entity the other has used the name to refer to and you now also will use the name as referring to that entity.

In addition to these two ways in which recipients can take uses of unfamiliar names there is a third one, where the interpreter simply does not know whether or not the name has a real referent and shelves this question for later. This third type of case is not uncommon either, and it is like the other two that in all three the communication may be the induction of the interpreter into the use of the name – make him part of that part of the speech community where this command is present.

How should MSDRT deal with such acquisitions of non-referring names? Intuitively the task is on the one hand to capture what these different case have in common, and on the other hand to capture the differences. The formal device we will use is a generalization of the one proposed for the use of names in fiction in (Kamp 2021*b*): The command by some agent A of a fictional name N involves, like the command of names for real things, a labeled entity representation of sorts. Such entity representations also feature anchors that are a bit like the vicarious anchors we have been speaking of so far; both kinds of anchors link the entity representations of which they are part with communicative acts by others from whom A got the name or that he takes to use the name in the same way that he does once he has acquired it. But there must be a feature of these new anchors that distinguishes them from vicarious anchors that are taken as witnesses to events of real reference. In (Kamp 2021*b*) it is assumed that entity representations for names belonging to different fictions (i.e. to different works of fiction or different mythologies) are marked accordingly. (Cases where an agent A learns a fictional name from some other agent, realizes it must be a name from some work of fiction or other, but doesn't know yet from which.) Here, where reference to entities belonging to established works of fiction and mythology are some but not all of our targets, this way of marking the relevant differences does not seem optimal. Instead we assume that these differences are marked on the 'anchors' that result from interpretations of uses of the name. (As before, we assume that each such act produces such an 'anchor', both when the name is heard or read for the first time and when the interpreter already has an entity representation for its use.)

Common between the various 'anchors' is that they all establish or reinforce links that connect the interpreter with those whose name uses he interprets. The differences between them are, to repeat, features associated with them. At this point we will do with just three features here: +real, -real and ?real. +real is for the vicarious anchors spoken of before the present section; it signals that the agent takes the entity representation to represent a real thing. -real signals the opposite: that the entity representation is for some fictional entity or for some illusory thing that is taken for real by others. And ?real marks anchors of entity representations of which one isn't sure whether they represent something real or not.

Two things remain to be decided. The first is a name for these new anchors. We propose, for lack of a better idea, to use the term *vicarious anchor* for all of them, thus extending the use that we have been making of this term hitherto. By adopting an entity representation with any such anchor, or adding any such anchor to an entity representation one already has, the interpreter commits himself to using the name in the same way as the speaker he is interpreting, whenever this entity representation is his basis for using the name.

The second decision is where to put the new features +real, -real and ?real in the representations. This is a question about our descriptions of entity representations as parts of our MSDs; in other words, about the form of ERs. Proposal: Tag the feature on as a fourth component.

Thus, instead of the ER $\left\langle [ENT,p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p,Paris) \\ \hline \end{array}, \mathcal{K}_p \right\rangle$ in (2.53) we now have: $$\left\langle [ENT, p], \begin{array}{|c|}\hline \\ \hline \text{city}(p) \\ \text{Named}(p, Paris) \\ \hline \end{array}, \mathcal{K}_p, +real \right\rangle$$

Furthermore we make '+real' into the default feature, which is to be understood when no feature is mentioned. In this way the old notation can still be used, with the same meaning.

Descriptive Anchors

Besides the anchors considered so far, which witness causal relations that ground the reference of their entity representations, we also allow for *descriptive anchors*. These attribute to the represented entity properties that it is supposed to satisfy uniquely. That is, a descriptive anchor of an entity representation takes the form of one or more predications of its distinguished dref. If there is a unique satisfier of this set of predications, in the world w and at the time t where and when this anchor is adopted, then that is, for all worlds and all times, the entity representation's referent. If there is no unique satisfier of these predications in w at t, then the entity representation is defective and represents nothing. Classical examples of entity representations with descriptive anchors are the introductions by Le Verrier of the names *Neptune* and *Vulcan* for the planet that he took to be the cause of certain previously unexplained aberrations in the orbit of Uranus and the one he took to be the cause of certain previously unexplained aberrations in the movement of the perihelium of Mercury, as told by Kripke in *Naming and Necessity*. According to the MSDRT account of this option for introducing names the result for the one who makes use of it is an entity representation labeled with the name and with an anchor set whose only member is a descriptive anchor that predicates the chosen definition of the distinguished dref. When the descriptive anchor is uniquely satisfied, then its unique satisfier is the referent represented by the entity representation. If it is not then there is no referent. *Neptune* is an example of the first possibility, *Vulcan* an example of the second.

For someone who makes use of the definition option for introducing a proper name N the initial result will be an N-labeled entity representation with the corresponding descriptive anchor as its only anchor. And so it will be for those who are privy to the introduction of the name. But it may not be that all are convinced that the definition is uniquely satisfied, so among the agents that have such an entity representation there may be some whose entity representations bear the feature ?real (and perhaps even -real) rather than +real. And as time goes on, these features may change as more becomes known about the relevant facts. As regards *Neptune* and *Vulcan* things notoriously went in opposite directions. In the case of the first there has been convergence on the feature +real, in that of *Vulcan* it has been on -real. (And as a consequence, *Neptune*-labeled entity representations now belong to the entity libraries of pretty much all reasonably educated people, whereas *Vulcan*-labeled entity representations are restricted to a comparatively small group of historians of science and philosophers.)

One can't help feeling that there is something provisional about many entity representations that are the initial results of such name definitions – not all of them, surely[2], but it is for the initial *Neptune*- and *Vulcan*-labeled representations, where the definition is tied to an as of yet unconfirmed scientific hypothesis. It is only when the hypothesis is confirmed, for instance by actual observation of the entity that is presumed to be represented by the entity representation, that one feels one is on the solid ground that proper entity representations require. At that point entity representations labeled with the name will have larger anchor sets, which have perceptual and vicarious anchors as well the initial descriptive anchor. (When the hypothesis is disconfirmed, then the name and the entity representations labeled by it will disappear, except for those who see such scientific mishaps as worthwhile objects of study – chapters in the history of science which may reveal interesting aspects of how the sciences work.)[3]

[2]In mathematics it is common practice to introduce names in this way when it is clear, or has been proved, that the definition is sound (i.e. that there is a unique satisfier). Think for instance of the definition of $\aleph_1$ as the first non-denumerable cardinal and countless other such definitions of names for mathematical objects.

[3]In earlier work the definition of a name and its bearer that is treated here as a type of anchor was located in the second, descriptive component of the ER. That is a natural place for it insofar as it is descriptive information about the referent and that is conceived as essential to its identity. Problematic about this was the assumption that this information in the descriptive component of the ER can be recognized to play its defining function from the fact that the anchor set is empty. That is fine when the anchor sets *is* empty. But that won't necessarily be the case. For as we will see, ERs that are introduced by definition originally can later acquire anchors of the kinds discussed in earlier sections of the present document. It is true that when those anchors then tie the ER securely to its referent, the status of the defining information that the ER kicked off with remains and it ought to be possible to recognize this from the form of the ER also at this later stage. This problem doesn't arise when the definition is made into an anchor, provided the forms of the different types of anchors always shows what kinds of anchors they are. Because of the special forms of perceptual and vicarious anchors this is, as far as I can see, always

3.1 Entity Representations as Individual Concepts

One of the uses that our present version of MSDRT makes of entity representations is best described as that of an *individual concept.* This kind of entity representation is needed in the context of desire, imagination, intention and planning – whenever agents conceive of things they want to find or create or that they hope or fear will come about or cross their paths. The representations involved in such mental activities are representations of entities, not propositional contents, thus their descriptions as parts of MSDs should be ERs of some sort. But they are, as just noted, not specific. They do not represent particular entities that belong to the real world. Nor do they represent characters of established mythologies or works of fiction, though they may develop local, semi-private narratives of their own. In MSDRT the term used to refer to such entity representations is *individual concept*, since that is for all practical purposes what they are. What is specific to MSDRT, though, is that (i) individual concepts are mental constituents and (ii) that's they can be shared between agents, as a result of communications between them. We will assume that non-reference-providing vicarious anchors create the links that are basis for such sharing.

While individual concepts are not representations of existing things, they can turn into such representations. (This is one aspect in which they differ from entity representations for mythological and other fictional characters.) Individual concepts are waiting to be instantiated, so to speak, by that which one wants to create or hopes to encounter. Some of them have to wait forever, and may disappear at some point from our mental states, through exhaustion or forgetfulness. But others are realized eventually and at that point change into entity representations of the kind discussed hitherto, representations of real entities. Such transitions from individual concepts to representations of entities with an independent existence in the real world is an important feature of the ways in which we interact with each other and with the physical world around us. (It is one of the insightful observations from (Ludwig 2016) about the transition from prior intentions to intentions-in-action.)

An example going back several decades (Kamp 1999) involves the following utterance from S to H.

(3.1) S to H: You remember that house we wanted to build on the Cornish

the case. If not, further notation would be needed to mark the different types of anchors.

coast twenty years ago? Well, last year we finally built it. You must come and see it.

The first and main point about S's words is the way she talks about the house then and now as one and the same thing. At the time she is asking H to remember the house was just a project. But as she says, that project was realized eventually and as a result there is now a real house that H can come and look at. That the project then and the concrete house today are treated as one and the same is strongly suggested by S's use of the pronoun *it* in the second and the third sentence.

Though (3.1) does not say so explicitly, the scenario it suggests is something like this: Twenty years ago S and one or more others – let us assume it was just her husband – formed the plan to build a house on the Cornish coast and at the time they talked about this to H, a friend. In her first sentence S appeals to his memory of those conversations twenty years earlier and the planned house that was their subject. And it is against the background of that memory that she then expresses the information that is new to him: the assertion about the building of the house and the invitation to come and have a look at it.

The MSDRT-inspired analysis of this is that at the time S and her husband formed a plan about building a house for which they each established, as part of their planning, an individual concept in the sense of MSDRT, i.e. as entity representations. Those entity representations were shared via (non-reference-establishing) vicarious anchors, added to their respective representations each time the other referred to the plan for the house. And when they then told H about the plan, he too formed an individual concept for the house, vicariously anchored to their individual concepts. The purpose of S's first sentence is to make H retrieve his individual concept for the house. If H does remember, by retrieving his old representation for the house, or accommodates by creating one afresh, then he can interpret S's uses of *it* in the next two sentences by using the distinguished dref of this individual concept as its antecedent.

We start a more formal description of this case by assuming that the relevant part of H's mental state as a result of his processing the first sentence of (3.1) is the individual concept displayed as the one relevant part of his current mental state in (3.2.a,b,c). ((3.2.a) is the official notation, (3.2.b) the abbreviated notation introduced in Ch. 2, Section 3. Furthermore, (3.2.c) is the relevant part of H's mental state after retrieval or accommodation of the individual concept. We will stick with the format in (3.2.c) for the remainder of our discussion of (3.1).)[4]

(3.2)

a.

$$
\begin{array}{|c|}
\hline
s \quad t \quad h \\
\hline
t = n \quad t \subseteq s \quad \text{person}(h) \\
s\colon \mathit{Att} \left(h, \left\{ \begin{array}{l} \left\langle [\mathit{ENT}, s_H]_{h,t}, \begin{array}{|c|}\hline \\ \hline \{\text{person}(s_H)\} \cup K_{s,H} \\ \hline \end{array}, \mathcal{K}_{s,H} \cup \left\{ \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\colon \text{see}(i,s_H) \\ \hline \end{array} \right\}, +\mathit{real} \right\rangle \\ \left\langle [\mathit{ENT}, m_H]_{h,t}, \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \cup K_{m.H} \\ \hline \end{array}, \mathcal{K}_{m,H} \cup \left\{ \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\colon \text{ref}^-(s_H,\tau,m_H) \\ \hline \end{array} \right\}, -\mathit{real} \right\rangle \end{array} \right\}, \emptyset \right) \\
\hline
\end{array},
$$

where τ is the noun phrase *that house we wanted to build on the Cornish coast twenty years ago.*

b.

$$
\begin{array}{|c|}
\hline
t \quad h \\
\hline
t = n \quad \text{person}(h) \\
\left\{ \begin{array}{l} \left\langle [\mathit{ENT}, s_H]_{h,t}, \begin{array}{|c|}\hline \\ \hline \{\text{person}(s_H)\} \cup K_{s,H} \\ \hline \end{array}, \mathcal{K}_{s,H} \cup \left\{ \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\colon \text{see}(i,s_H) \\ \hline \end{array} \right\}, +\mathit{real} \right\rangle \\ \left\langle [\mathit{ENT}, m_H]_{h,t}, \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \cup K_{m,H} \\ \hline \end{array}, \mathcal{K}_{m,H} \cup \left\{ \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\colon \text{ref}^-(s_H,\tau,m_H) \\ \hline \end{array} \right\}, -\mathit{real} \right\rangle \end{array} \right\} \\
\hline
\end{array}
$$

[4]Using the letter m for the distinguished dref of the house concept may be suboptimal. But h is misleading because it might suggest the hearer H. m is for French ‘maison’.

c.

$$\left\{\begin{array}{l}
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \{\text{person}(s_H)\} \cup K_{s.H} \\ \hline\end{array}, \mathcal{K}_{s,H} \cup \left\{\begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: see}(i,s_H) \\ \hline\end{array}\right\} \right\rangle, +real \\
\\
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \cup K_{m.H} \\ \hline\end{array}, \mathcal{K}_{m,H} \cup \left\{\begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}^{-}(s_H,\tau,m_H) \\ \hline\end{array}\right\}, -real \right\rangle
\end{array}\right\}$$

The part of H's mental state described in (3.2.a,b) consists of (i) H's entity representation for S and (ii) the entity representation for S's planned house that he has either retrieved or accommodated. New is the notation used in the overtly presented vicarious anchor in the Entity Representation for the planned house. The predicate 'ref^{-}' is like the predicate 'ref' in (2.51) in that it links the Entity Representation R ER in whose anchor set it occurs to a 'reference' made in the utterance that is being interpreted. But 'ref^{-}' differs in that it does not assume that the speaker was referring to a real entity, and thus it does not make ER a representation of what the speaker has been referring to (since she is not taken to have made a reference to some real entity). In what is coming we will often have to distinguish between reference-conferring vicarious anchors and those that only function as witnesses to a link between producer and interpreter.

Related to this is the implicit assumption that none of the anchors in $\mathcal{K}_{m,H}$ is of the reference-witnessing kind: This ER describes an entity representation of H that he does not take to represent a real entity.

To interpret the second sentence of (3.1), H makes use of his ER for the planned house that he has just been made to retrieve or re-establish. What must happen as part of his interpretation of this second sentence is of central importance to the use we are going to make of MSDRT. The natural, and in the present context unavoidable, understanding of the second sentence is that the building it described was the execution of the project of which S has just reminded H with her first sentence. We will have much more to say about plan structure and its role in plan execution later on. For now we focus just on what such a 'plan execution' interpretation of the second sentence entails for H's entity representation of the planned house.

In principle there are two positions that could be chosen here. One would be to adopt a new entity representation for the built house and a belief to the effect that the entity it represents is the result of execution of the project to realize the individual concept represented in (3.2). The other is to assume that the house was there all along (for the past twenty years) albeit for some of that time only as a concept. In view of what we have said about the form of (3.1) it will be no surprise that we go for the second option.

But either way whatever is needed at this point is a way of reflecting about the life spans of mental state constituents and parts thereof. This is true both about our propositional attitudes and about our entity representations. As regards the latter: among the thoughts that agents have concerning their entity representations are thoughts about how long they have known the represented entities, what they knew about them before they met them in the flesh, how eager they were, and when, to meet them personally, and so on and so forth. This is yet another reason why we need a way to identify MSD constituents as temporal continuants, an issue raised at the end of the section 'Abbreviations' (Section 3 of Ch. 2). Once more we refer to Ch. 4, Section 4.1 for details.[5]

Part of the information conveyed to H by S's second sentence is that there was an event in the past that consisted of building the house that thus far had existed as the individual concept referred to in the first sentence. For H this is the signal to modify his current ER for the house from -real to +real. This switch is the result from his interpretation of the pronoun *it* in

[5] In other cases, when we want to express that an entity representation has a certain property at a certain time and where it is therefore necessary to refer to the entity representation as argument (as distinct from referring to the entity it represents), another device will be needed. The device we will adopt is that of providing the distinguished dref of the ER with a hat. Thus when A has an entity representation that is ascribed to her as part of her mental state in some w at some t and this entity representation is described by the ER '$\langle [ENT, x]_{a,t}, K, \mathcal{K} \rangle$', then $\hat{x}$ will serve as argument term denoting this entity representation in predications that attribute properties to the entity representation at this or some other world-time pair.
The same issue also arises for propositional attitudes. They too are typically part of an agent's mental state over some period of time. But in the MSDs we have been using the descriptions of propositional attitudes do not come with unique identifiers. We will introduce unique identifiers for Propositional Attitudes when we will need them.

the second sentence. On the one hand the resolution of *it* is provided by the second of the two ERs in (3.2.c). But the displayed vicarious anchor of the second ER is a witness to S's use of *that house we wanted to build on the Cornish Coast twenty years ago* to refer to the entity represented by an entity representation she has for the planned house. Her own resolution of the pronoun in the second sentence of her utterance would be resolved to this entity representation of hers. But on the other hand the predication expressed by the second sentence indicates that this entity representation must now represent a real house. H takes S to have used *it* to refer to the entity represented by this entity representation of hers, and that will induce the addition of yet another vicarious anchor to his entity representation. But this time that will be a vicarious anchor which confers proper representation on his entity representation. Recall that a general form for vicarious anchors was presented as (2.49) in Ch. 2, section 4. We repeat:

(2.49)

e
$e \prec n$ e: ref(s,γ,α)

In the case before us H's vicarious anchor is an instance of this, with *it* for γ and m_H for α:

(3.3)

e'
$e' \prec n$ e': ref(s_H,it,m_H)

By adding this vicarious anchor to his entity representation for the house H turns the representation into one that does represent an existing object, one that is described by a +real ER in the MSD for the relevant part of his MSD at this point.

All this is part of H's interpretation of the propositional content of the second sentence. When the adverb *finally* is ignored[6], then a content DRS for this sentence can be constructed using familiar DRT rules, most of which we have encountered. A comment on the lexical entry for the verb of creation *build*

[6]It is intuitively more or less clear what *finally* contributes to the sentence: that it took a long time for the event of building the house to happen. A proper account of the contributions that such 'commenting' adverbs make to sentence content or meaning, is a notoriously difficult problem. We are unaware of any solutions to it that are formally explicit enough to allow for direct incorporation into a DRS construction algorithm.

follows below. We assume once more that H accepts what S says as true, adding its content representation to his mental state as a belief. (3.4) is the MSD for the relevant part of his mental state as resulting from his dealing with the first two sentences of (3.1).

(3.4)

$$
\left\{
\begin{array}{l}
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \{\text{person}(s_H)\} \cup K_{s.H} \\ \hline \end{array}, \mathcal{K}_{s,H} \cup \left\{ \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: see}(i,s_H) \\ \hline \end{array} \right\}, +real \right\rangle \\[3em]
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \cup K_{m,H} \\ \hline \end{array}, \mathcal{K}_{m,H} \cup \left\{ \begin{array}{c} \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}^{-}(s_H,\tau,m_H) \\ \hline \end{array}, \\[1em] \begin{array}{|c|}\hline e' \\ \hline e' \prec n \\ e'\text{: ref}(s_H,\tau,m_H) \\ \hline \end{array} \end{array} \right\}, +real \right\rangle \\[3em]
\left\langle BEL, \begin{array}{|c|}\hline t'' \quad e'' \quad X \quad y \\ \hline t'' \prec n \quad \text{c-year}(y) \quad t'' \subseteq y \quad e'' \subseteq t'' \\ s_H \in X \quad X \supset s_H \\[1em] \begin{array}{|c|}\hline y' \\ \hline \text{c-year}(y') \quad n \subseteq y' \quad y \supset\subset y' \\ \hline \end{array} \\[1em] e''\text{: build'}(X,m_H) \\ \hline \end{array} \right\rangle
\end{array}
\right\}
$$

Legenda Part of the content DRS for the belief in (3.4) has to do with the contribution made by *last year*. *Last year* is analyzed as the calendar year y that abuts the current calendar year y' on the left. (The predicate 'calendar year' is represented as 'c-year'.) The sub-DRS involving y' is another example of a DRS Condition without an overt operator, and an existential semantics, which permits it to be merged with the DRS to whose Condition Set it belongs.

What remains is the interpretation of S's third sentence. But nothing new, as far as we can see, can be learned from going through this for our present purposes. A detailed look at this third sentence would be interesting because of the Speech Act related question it raises – that it functions as an invitation, while being in the form of a modal declarative. But this belongs to another chapter of MSDRT: its application to the Theory of Speech Act Types, and that chapter still needs to be written.
[end Legenda]

Build is a *verb of creation*: it describes events at the end of which something exists, as the causal result of that event, that didn't exist when the event began. In virtue of this H can infer that the house now represented by his ER with distinguished dref m_H exists at the end of the event e'' in (3.4) and that it didn't exist before e'' began. In other words, it would be fully legitimate for H to adopt the further belief that this is so. This belief can be expressed, moreover, by adopting an existence predicate EXIST that is true of an object **d** from an IH model M, in a world w of M at a time t of M iff **d** belongs to the Universe of the extensional model $M_{<w,t>}$. With this predicate the content of this belief can be represented as in the DRS (3.5).

$$
(3.5)\quad
\begin{array}{|c|}
\hline
s_1 \quad s_2 \\
\hline
s_1 \supset\subset e'' \supset\subset s_2 \qquad n \subseteq s_2 \\
\\
s_1\text{: } \neg\ \text{EXIST}(m_H) \qquad s_2\text{: EXIST}(m_H) \\
\hline
\end{array}
$$

(Here 's_1: ¬ EXIST(m_H)' is a shorthand for $s_1\text{: } \neg \begin{array}{|c|} \hline s_3 \\ \hline s_3 \subseteq s_1 \\ s_3\text{: EXIST}(m_H) \\ \hline \end{array}$.)

One moral to draw from this example is how the reference-related and the communication-related aspects of vicarious anchors can be intertwined. S's mention of the house she and her husband had been wanting to build for many years leads H to reactivate or renew the entity representation that he shared with them many years ago. That was, like theirs at the time, one with the status of an individual concept. But then, when S refers back to her entity representation for the house by means of the anaphoric pronoun *it* while placing *it* in an argument position that selects for real physical objects, and H is made to resolve *it* to his entity representation of the house, this leads him to the conclusion that the individual concept has been instantiated in the meantime. The vicarious anchor that he adds to his entity representation for the house in response to S's use of *it* in this sentence, and that has, as distinct form the first vicarious anchor, the function anchoring representation to a real referent, functions as a kind of correction: the entity representation is the representation of a real referent and not just a mere, uninstantiated, individual concept.

The example is also an illustration of how shared individual concepts that coincided fully or largely when they were first conceived and came to be shared, can come apart subsequently, when the sharers no longer interact with each other. That since the time when S and H interacted, S's concept of the planned house has mutated into the representation of an existing object, whereas H's has not changed in this way is just one example of this. But it may be taken for granted that the drifting apart of the two concepts involved much more than this. As S and her husband started to implement their project, fixing the place where the house was to be built and acquiring the land, then getting an architect to make the plans and dealing with details of the house that they had never thought of before, meant that their concepts of what they were going to build and then were building changed all the time and, for all the story tells us, sometimes drastically. And through all this H's concept remained entirely unaffected.

Nothing of this is in any way remarkable or surprising. We lose track of the people and things that we once knew while others were kept more in the loop, and then ever so often they can tell us about what happened after we did lose touch. That is the function and charm of gossip. And what goes for our information about real people and things, goes even more so for the plans that others once shared with us and of which do not know what came of them. Still, it is the continued sharing of those now widely diverging entity representations that makes it possible, and often so very easy, to reestablish contact about what they represent.

A central feature of the classical notion of an individual concept is uniqueness: In intensional semantics an individual concept is a function from possible worlds to individuals: one, unique individual per possible world. One might also recognize partial individual concepts, partial finctions from the set of worlds to individuals, but that preserves the notion of (at most) one individual for any one world. For MSDRT's individual concepts uniqueness is guaranteed in a different way. With an individual concept comes a notion of its canonical instantiation. For individual concepts that are formed in the course of planning or the formation of intentions this canonical instantiation gets produced when the plan or intention is successfully executed. And successful execution of any given plan or intention can occur only once; as soon as that happens the plan or intention will be laid ad acta, as something that has been successfully dealt with. But of course this is a notion of uniqueness in the sense of at most one; for intentions often remain unsatisfied and many plans are never get carried out.

Uniqueness of individual concepts in MSDRT may also be guaranteed in other ways, for instance by the first instantiation of it that crosses the path of the one who has the concept (or of someone belonging to the group of those who share it). What other forms of unique instantiation there may be needs further exploration.

MSDRT makes it possible to describe this case in more detail, tracking the assumptions made by the participants one level further down. The way in which S phrases her message to H in (3.1) reveals her take on what they share and what they don't share. And the way her message to H is structured is designed to exploit what they share to remove the difference. At the point when she starts her utterance, S correctly assumes that at the time that her first sentence is trying to remind H of the two of them shared the individual concept of the house that she and her husband had formed the plan or desire to build. That is, she knows that at that earlier time her own entity representation she now has for the house that now exists still was an uninstantiated individual concept, and that H too had such an individual concept and that these two entity representations were linked – by perhaps a considerable number of vicarious anchors on either side, stemming from the conversations they had been having about the project. What can also be inferred from her choice of words is that she assumes that for H the entity representation he adopted at the time as part of their conversations about the project established by their conversations then still has the status (from H's own perspective) of an uninstantiated individual concept, something that is no longer in tune with the way things are, as the house has now been built.

To describe these various assumptions that motivate S to express herself the way she does at the time when she utters (3.1) we rely on the unique identification of entity representations by their distinguished drefs.

(3.6)

$$
\left\{
\begin{array}{l}
\left\langle [ENT,h], \begin{array}{|c|}\hline s_0 \\ \hline n \subseteq s_0 \quad \text{person}(h) \\ s_0\text{: addressee}(h,i) \\ \hline \end{array}, \left\{ \begin{array}{c} \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: see}(i,h) \\ \hline \end{array} \\ \\ \begin{array}{|c|}\hline s_2 \\ \hline n \subseteq s_2 \\ s_2\text{: talk-to}(i,h) \\ \hline \end{array} \end{array} \right\}, +real \right\rangle \\
\\
\left\langle [ENT,m], \begin{array}{|c|}\hline \\ \hline \{\text{house}(m)\} \cup K_m \\ \hline \end{array}, \mathcal{K}_m , +real \right\rangle \\
\\
\left\langle BEL, \begin{array}{|c|}\hline t \quad s_4 \\ \hline t \prec n \quad |(t,n)|_{years} = 20 \quad t \subseteq s_4 \\
\left\langle [ENT,m]_{i,t}, \begin{array}{|c|}\hline \\ \hline \{\text{house}(m)\} \cup K_{m,t} \\ \hline \end{array}, \mathcal{K}_{m,t} , -real \right\rangle \\
\left\langle [ENT,m_H]_{h,t}, \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \\ \cup K_{m,H} \\ \hline \end{array}, \mathcal{K}_{m,H}, -real \right\rangle \\
s_4\text{: DLink}(\hat{m},\hat{m}_H) \\
\left\langle [ENT,m_H]_{h,n}, \begin{array}{|c|}\hline \\ \hline \{\text{house}(m_H)\} \\ \cup K_{m,H} \\ \hline \end{array}, \mathcal{K}_{m,H}, -real \right\rangle \\
\\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

Legenda

1. (3.6) is an MSD for S's mental state at the utterance time of (3.1). It

consists of (i) an ER for S's representation of her interlocutor H; (ii) an ER for her current representation of the house she and her husband have built on the Cornish coast; and (iii) a belief to the effect that (a) there was a time 20 years ago when both she and H had individual concepts for a house that she and her husband were planning to build on the Cornish coast and that these entity representations were 'D-Linked' at the time (see below), and (b) that H still has his individual concept for the house.

2. As regards the representation of the belief, observe that much of this representation is now hidden. Each of the three ER-like constituents in its content representation is a set of DRS Conditions the main one of which is an *Att*-predication, while the other conditions locate the state argument of that *Att*-predication in time. (It is a useful, if somewhat boring, exercise to write out this content-DRS in official, unabbreviated notation.)

3. The belief content is to the effect that H's current entity representation still has the status of an uninstantiated individual concept (indicated by the feature '-real'), whereas S's own entity representation for the house has evolved from a mere individual concept to a fully fledged entity representation of an existing entity (switch from '-real' to '+real').

4. The most important new feature of (3.6) is the Condition 's_4: DLink($\hat{m}$,$\hat{m}_H$)'. This Condition says that at t S's and H's individual concepts for the house were *D-Linked*. D-Linking is between an entity representation of an agent A and an entity representation of an agent B. It is often the effect of B adding a vicarious anchor to his entity representation as witness of A having used a definite noun phrase to refer to what is represented by her entity representation and of B interpreting the noun phrase in this sense. But more needs to be said about it. Some of that follows right here.

end Legenda

D-Links and Linking

D-Linking is a relation between Entity Representations. In that respect it is like the relation that gets established by vicarious anchoring, with vicarious anchors as we have described them: as witnesses that an agent adds to an ER of references by others to the entity that he takes his ER to represent. This presupposes that the agent's ER does represent an entity, and also that the definite noun phrase whose utterance gives rise to the anchor properly refers, and thus, in case the utterer used the definite noun phrase to refer to the entity represented by an ER of hers, that ER was an entity representing one as well. When these presuppositions are fulfilled, then the vicarious anchor establishes that the two ERs are coreferential. We can now add that the vicarious anchor also establishes a D-Link between the two ERs in this case. But D-Linking is a more general notion, which also applies when one or both of the two ERs do not represent an entity – and in particular, when, in the terminology introduced in this section, they bear the feature '-real', like the ERs functioning as individual concepts of which the agent assumes that they are uninstantiated.

It is important here to distinguish between cases of unintended reference failure, when the recipient of a definite noun phrase erroneously takes it to refer and interprets it either by using an ER he already has or by accommodating a new ER for the purpose, and cases where no error on the part of the interpreter is involved. First consider the case where the interpreter is wrong to think that the definite noun phrases used by the speaker properly refers. According to what we have been saying about vicarious anchoring, his interpretation of the speaker's use of the definite noun phrase as properly referring to an entity for which she has an ER take the form of adding a vicarious anchor to his old or new ER, as a witness to this interpretation of the definite noun phrase. When the interpreter uses an ER he already has, there are two possibilities: (i) it may be one that does properly represent, but that now gets corrupted by a vicarious anchor that is defective because of its false assumption that the speaker had a properly representing ER for the referent of the definite noun phrase; or (ii) contrary to what the interpreter assumes, his ER doesn't properly refer. In this case the added vicarious anchor doesn't make things worse as far as proper representation goes, but it doesn't make them any better either. The case where the interpreter accommodates a new ER is like (ii).

The possibilities of the last paragraph can be considered to have been covered by what has been said about vicarious anchoring up to now. Mistaking a definite noun phrase use to be one that properly refers can be regarded as similar to the optical illusion cases mentioned in Ch. 2, Section 1. But there are also cases involving ERs that fail to properly represent, but where the agent is aware of this. ERs that serve as uninstantiated individual concepts are one type of example of this. (In the next section we will encounter another type.) We assume, as part of our treatment of coordinated individual concepts, that when 20 years ago S and H talked about the house that S and her husband intended to build, they coordinated their respective individual concepts for that house each time that one of them mentioned it. We also assume that this coordination takes a similar form to that of adding vicarious anchors: the recipient of the mention adds a witness of their communication about the house to the third component of his individual concept ER. We are reluctant to refer to this third component as the ER's 'anchor set', since the members of this set aren't anchors in the sense so far considered: they aren't witnesses to coreference, and aren' intended as that. They are just records of the recipient's belief that the current mention is to what the two of them have coordinated concept representations for. We refer to such witnesses, which are added to the third components of individual concept ERs, also as *links*. Links, in this sense, re alternative ways in which ERs can get D-Linked.

On page 156) we proposed (3.3) as vicarious anchor that H adds to his individual concept ER for the house. We added there – making explicit what we take to have been understood all along – that such anchors commit the agent of the ER to which they are added to the ER being one that properly refers. Links do not carry such a commitment, so (3.3) cannot be used as representational form for them. The form we propose instead has 'men' (for 'mention') in lieu of 'ref':

$$(3.7)\quad \begin{array}{|c|} \hline e' \\ \hline e' \prec n \\ e'\text{: men}(s_H,\gamma,m_H) \\ \hline \end{array}$$

In (3.6) the D-Link between the ERs with distinguished drefs m and m_H was represented as one of the Conditions that make up the belief that S is assumed to have at the time when she says (3.1) to H. That is a legitimate way of coding this information, but it does not fit optimally with other aspects of the notation we have adopted as we went along. In the full representations of MSDRT, in which propositional attitudes are constituents of MSDs that occur in the 3rd slot of some *Att*-predication, the 4th slot of that predication will be filled with a set of pairs $<\alpha,\alpha'>$, where α is the distinguished dref of an ER that is a constituent of the MSD occurring in 3rd position and α' the distinguished dref of an ER that occurs elsewhere in the representation, outside the given *Att*-predication. Often the justification of such a link-pair is that the first ER contains a vicarious anchor that connects it with the second ER. But the connection need not be this direct. In general $<\alpha,\alpha'>$ means that the two ERs are connected by a chain of direct links, between ERs of possibly many different agents. Such multi-link chains still imply that there is a commitment for the agent of the first ER that it refer to the same entity as the second ER.

Linking between individual concept ERs may be similarly direct or indirect. In our example of the house on the Cornish coast the link is a direct one; but indirect links can come about in the same way as the indirect links that carry coreference commitments. It seems natural to represent direct and indirect D-links also within the link sets of *Att*-predications, for after all they too are connections between ERs of the predication's MSD and ERs outside the predication. But since D-links do not carry coreference commitments, they should be given a different form. Here are two options that readily come to mind. The first uses the predicate *DLink*, with the two distinguished drefs as arguments; so in the present case this will be 'DLink(m,m_H)'. The second option takes the form of an ordered pair, but now consisting of two terms denoting the ERs themselves, as distinct form their distinguished drefs. Given our convention to use hatted distinguished drefs as terms denoting their ERs, this would in the present case look like this: '$<\hat{m},\hat{m}_H>$'. We will allow for either option. The first has the advantage of making visually prominent that we are dealing with a D-Link rather than a referential link; the second fits in more smoothly with the notation for referential links we have been using. To use either of these options we need the notation in which the first ER is part of an MSD that occurs as 3rd argument of an *Att*-predication. It is awkward to illustrate this with alternative version of (3.6). But we will use the new notation in the next section and again in Ch. 3, Section 2.

The discussion above of referential links and D-Links suggests that the former entail the latter. D-Links are about an intersubjective relation between the mental states of different agents. Referential links are in the first instance about how an agent's mental state is related to the 'world'– to what the contents of the Propositional Attitudes in her state are about – and only as a kind of side effect, you might say, about how her state is related to one or more other mental states. Referential links do two things and Dlinks do just one of those.

This picture is correct for the mental acts that link an agent's mental state to that of another. When an interpreter adds a vicarious anchor to one of his ERs, he does two things: link his ER to that of the speaker whose utterance he interprets and thereby also fix or reinforce the relation between his ER and the entity it is meant to represent. When he adds a D-Link to his ER, all he does is establish a link with an ER of the speaker. Vicarious anchoring is D-Linking for the sake of reference coordination. There is entailment in one direction but not in general in the other. By the same token, when an indirect link between two ERs has come about through a chain of vicarious anchorings, then it entails a D-Link between the two ERs. And once again, not in general the other way round.

But this does not mean that all cases of coreferential ERs are cases of intersubjective linking. You are in Room 2 and you know that A is in Room 2 and B is in Room 3, but A and B do not know that of each other. Then there is a very loud bang, which all three of you must have heard. So you infer that both A and B now have ERs for the bang and that each of those is coreferential with your own ER for the bang. There is then a coreference chain between A and B, with you in the middle. But there is no intersubjective connection between the ERs of A and B. D-Links between ERs of different agents can only come about, we hereby stipulate, when at least one of the ERs contains a D-Link in its third component.

One more general point about D-Links. So far we have operated on the assumption that direct links between entity representations of utterers and interpreters result from the anchors (or D-Links) that interpreter adds to the third component of his ER. Nothing, we have thus far assumed, happens to the ER of the utterer. But that is unrealistic. In face-to-face communication interpreters react to the utterances they attend to with a multitude of signals, that they understand, do not quite understand, approve, doubt or disagree. That is an essential part of the cybernetics of conversation. Non-negative feedback, including absence of protest, will often serve the speaker as confirmation that her addressee is sharing her familiarity with the entities to which she refers, or that he accepts their existence on the strength of what she says. It seems a reasonable assumption that in such face-to-face situations the speaker records that she shares the ERs on which she has based the referring expressions she has used with co-referring ERs in the mind of her interpreter. In such situations the coordination of ERs can be expected to often be a two- rather than a one-sided affair.

In a more fine-grained version of MSDRT, which pays closer attention to the phenomenology of conversation, we might add machinery to represent the ways in which speakers' Entity Representations are modified in response to apparently successful references. But in the present Introduction the overload this would produce does not seem to serve a useful purpose.

So much for now about D-Linking and its relation to vicarious anchoring.

(3.6) has been our first explicit example of how speakers can reflect about the present and past mental states of their addressees and also about their own, and how such reflection can be the basis for the particular way they phrase their communications. As far as this last point goes, there is of course still much that (3.6) leaves unarticulated: S's awareness of the fact that she and H haven't been in touch since the time when the building of the house got under way and her inference from this that his representation for the house continues to be one that has the status of an uninstantiated individual concept. But at least we have the beginnings of the mutual assumptions that are made as a matter of course between joint planners and joint executors.

But a crucial feature of an account of why (3.1) has the form it has and what its communicative point is is in place with (3.6): S's assumption that H is not aware that the house they talked about 20 years ago was actually built at some later point in time. And that this discrepancy between his information and her own takes the form of him still having the same individual concept that he had when they last communicated about the house-building project whereas her individual concept has turned into the representation of an object in the physical world. The second sentence of (3.1), the one that conveys the new information that S conveys to H, serves to eliminate this difference between the two of them in that it endeavors to get H to also turn his individual concept into the entity representation of the entity that is represented by hers.

This also points to an important aspect of the sense in which vicarious anchors that do not establish or confirm representation of some independently identifiable entity nonetheless establish a form of *coreference*: Whatever will count as being the entity represented by the source's entity representation will ipso facto also count as represented by the entity representation of the interpreter who has added the vicarious anchor to it.

3.1.1 Santa

In our last example there was an initial discrepancy between D-Linked entity representations of the two participants in a verbal communication, in that the speaker knows that her entity representation represents an actual object whereas for the addressee his entity representation was still playing the part of an individual concept. But in this example that is just a transitory state of affairs, which S's second sentence is meant to put an end to (and does, as the story has been told here). But in the next example the discrepancy is stable, with the one whose entity representation is not assumed by her to properly represent acting so as to avoid detection by the other participant, who erroneously takes his entity representation to be the representation of a real thing.

The case is that of Mummy and Johnny who talk – now and more regularly – about Santa. Mummy knows, to the extent that such things can be known, that Santa doesn't exist, isn't a real person, but Johnny believes there is. Nevertheless, when they talk about Santa, using the name *Santa* or pronouns/demonstratives to refer back to earlier occurrences of the name, communication seems to work just fine, without qualification for Johnny and within the limits she has set herself for Mummy.

From an MSDRT-based perspective of what is going on in exchanges between Mummy and Johnny there is not much to be added to what has been said about verbal communication involving D-Linked n-labeled entity representations in the mental states of the participants involved. The only difference is that in this case the entity representations retain their respective status: Johnny has a *Santa*-labeled entity representation with the feature +real, but that is defective – given that it is +real, it should have a referent but it doesn't – and Mummy has a *Santa*-labeled entity representation with the feature -real that isn't defective – it doesn't have referent either, but it is not supposed to.

But these differences do not impede communication. As Santa gets mentioned by either Johnny or Mummy, the other's entity representation gets activated and its anchor set augmented. In this fashion the anchor sets of their *Santa*-labeled entity representations grow, sustaining their sense that they are talking about the same individual whenever either of them uses the name *Santa*.

These aspects of Mummy and Johnny talking about Santa are as expected. But the example raises another issue, which arises generally in connection with names for fictional entities and that we mention here in passing. Consider the following two sentences and assume that both are uttered by Mummy.

(3.8) a. Santa has a long white beard.

b. Santa is coming tomorrow.

When uttered by Mummy, (3.8.b) is an outright lie. A white lie arguably – how white may depend on your educational principles and on Johnny's age – but a lie no less: No one that could be the referent of Mummy's and Johnny's D-.Linked *Santa*-labeled entity representations will come tomorrow. Mummy knows that perfectly well, and if she says this nevertheless, it is as part of the big conspiracy to keep Johnny from wondering whether Santa really exists. (What would childhood be without such conspiracies?) But (3.8.a) is a different kind of statement. It won't be true in quite the way in which it was true that Charles Darwin or Leonid Tolstoj had a white beard. At least not by Mummy's count, since she doesn't think that Santa ever existed at all. Nevertheless there is something true about (3.8.a): According to received Santa mythology this is what Santa looks like; what stands out, in each and every reenactment of him is the striking contrast between his snow white beard and his bright red outfit. That makes (3.8.a) 'true in the story' as it is usually put; and 'true in the story' in the story so long as there are no competing criteria. The distinction between statements involving characters of fiction that are true in the sense of 'true in the story' and those whose truth conditions are fixed in other ways is a notoriously hard problem for the semantics of fiction.[7]

Much work on the use of names in fiction and metafiction tacitly assumes that everyone involved in the use of fictional names is aware that they are fictional. That assumption need not be right for all cases, but it must be largely right for 21-century largely secularized adults. Our Santa scenario is different in this respect. One of the participants, Johnny, gets things systematically wrong, with the connivance of his mother who avoids saying anything that might rock his conviction that Santa exists. Johnny has a *Santa*-labelled +real ER, and with each occasion when he hears someone say something about Santa, this will reinforce the anchor set for this ER with yet another faulty vicarious anchor. Johnny thinks that the other is using *Santa* to refer to a real person, and adds this assumption in the form of a new anchor to the anchor set of his *Santa*-labelled ER. But all those anchors are void, either because the user of *Santa* didn't use it as the name of a real individual or because the other person was like Johnny in believing that Santa exists, and has an equally ill-grounded *Santa*-labelled +real ER as Johnny himself.

The case of mummy and Johnny is one where we can put our discussion about vicarious anchors and D-Linking from the last section to work. (3.9.a) and are parts of the mental states of Johnny and mummy which consist of their respective entity representations for Santa together with representations for each other. Johnny's *Santa*-labeled ER is +real, mummy's is -real.

[7]For discussion of this problem in an MSDRT-based setting see (Kamp 2021*b*).

(3.9) a.
$$\left\{\begin{array}{l}
\left\langle [ENT, sc_j], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_j) \\ \text{Named}(sc_j, \textit{Santa}) \\ \hline \end{array}, \mathcal{K}_{sc,j}, +real \right\rangle \\
\\
\left\langle [ENT, m_j], \begin{array}{|c|}\hline \\ \hline \text{person}(m_j) \\ \text{Named}(m_j, \textit{Mummy}) \\ \hline \end{array}, \mathcal{K}_{m,j}, +real \right\rangle
\end{array}\right\}$$

b.
$$\left\{\begin{array}{l}
\left\langle [ENT, sc_j], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_j) \\ \text{Named}(sc_j, \textit{Santa}) \\ \hline \end{array}, \mathcal{K}_{sc,m}, -real \right\rangle \\
\\
\left\langle [ENT, j_m], \begin{array}{|c|}\hline \\ \hline \text{person}(j_m) \\ \text{Named}(j_m, \textit{Johnny}) \\ \hline \end{array}, \mathcal{K}_{j,m}, +real \right\rangle
\end{array}\right\}$$

When mummy uses *Santa* when talking to Johnny and Johnny has to understand what mummy means by *Santa*, Johnny will proceed, we claim, the way we have been proposing since the beginning of Ch. 2: He will add a vicarious anchor to his ER for Santa. That will be wrong of course, since Santa doesn't refer to an ordinary person, contrary to what he thinks. But that won't prevent him from doing this. The result is the update shown in (3.10.a) of his ER in (3.9.a).

When, on the other hand, Johnny uses *Santa* and his mother has to interpret this use, she won't add a vicarious anchor to her *Santa*-labeled ER. His use of *Santa* won't persuade that Santa was a real individual. But she knows that Johnny is trying to refer to the imaginary person for which she has her non-referring ER in (3.9.b). This acknowledgement will take the form of her adding a D-Link to the third component of her ER for Santa.

(3.10)

a.
$$\left\{\begin{array}{l}
\left\langle [ENT, sc_j], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_j) \\ \text{Named}(sc_j,Santa) \\ \hline\end{array}, \mathcal{K}_{sc,j} \cup \left\{ \begin{array}{|c|}\hline e_{utt,m} \\ \hline e_{utt,m} \prec n \\ e_{utt,m}\text{: ref}(m_j,S'a,sc_j) \\ \hline\end{array} \right\}, +real \right\rangle \\
\\
\left\langle [ENT, m_j], \begin{array}{|c|}\hline \\ \hline \text{person}(m_j) \\ \text{Named}(m_j,Mummy) \\ \hline\end{array}, \mathcal{K}_{m,j}, +real \right\rangle
\end{array}\right\}$$

b.
$$\left\{\begin{array}{l}
\left\langle [ENT, sc_m], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_m) \\ \text{Named}(sc_m,Santa) \\ \hline\end{array}, \mathcal{K}_{sc,m} \cup \left\{ \begin{array}{|c|}\hline e_{utt,j} \\ \hline e_{utt,j} \prec n \\ e_{utt,j}\text{: men}(j_m,S'a,sc_m) \\ \hline\end{array} \right\}, -real \right\rangle \\
\\
\left\langle [ENT, j_m], \begin{array}{|c|}\hline \\ \hline \text{person}(j_m) \\ \text{Named}(j_m,Johnny) \\ \hline\end{array}, \mathcal{K}_{j,m}, +real \right\rangle
\end{array}\right\}$$

This is as good a moment as any to make explicit what would be hard to refuse to acknowledge. If one agrees with our central tenet that agents have minds replete with entity representations and that they use these in the production and interpretation of referential expressions along the lines developed in MSDRT, then surely agents will also make assumptions about entity representations in the minds of those they communicate with. We have relied on this more than once already when we assumed agents to be in mental states that include beliefs about the mental states of others and where the latter contain ERs as well as Propositional Attitudes.

Our reason for drawing attention to this hardly controversial point here has to do with the way in which the operations of adding vicarious anchors and D-Links to ERs are related to the links and D-Links in the link sets of *Att*-predications. The MSDs in (3.10.a,b) present very small excerpts from the mental states of mummy and Johnny. In the light of our uncontroversial observation they can be extended with the beliefs that have been included in the MSDs in (3.11).

(3.11)

$$
\text{a.}\ \left\{ \begin{array}{l}
\left\langle [ENT, sc_j], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_j) \\ \text{Named}(sc_j, \mathit{Santa}) \\ \hline \end{array}, \mathcal{K}_{sc,j} \cup \left\{ \begin{array}{|c|}\hline e_{utt,m} \\ \hline e_{utt,m} \prec n \\ e_{utt,m}\text{: ref}(m_j, S'a, sc_j) \\ \hline \end{array} \right\}, +real \right\rangle \\
\\
\left\langle [ENT, m_j], \begin{array}{|c|}\hline \\ \hline \text{person}(m_j) \\ \text{Named}(m_j, \mathit{Mummy}) \\ \hline \end{array}, \mathcal{K}_{m,j}, +real \right\rangle \\
\\
\left\langle BEL, \begin{array}{|c|}\hline s_{m,bel} \\ \hline n \subseteq s_{m,bel} \\ s_{m,bel}{:}Att \left(m_j, \left\{ \begin{array}{l} < [ENT, sc_m]\ , \\ \begin{array}{|c|}\hline \\ \hline \text{person}(sc_m) \\ \text{Named}(sc_m, \mathit{Santa}) \\ \hline \end{array}, \\ \mathcal{K}_{sc,m} \cup \left\{ \begin{array}{|c|}\hline e_{utt,j} \\ \hline e_{utt,m} \prec n \\ e_{utt,m}{:}\mathbf{ref}(m_j, S'a, sc_j) \\ \hline \end{array} \right\}, \\ \qquad\qquad +real > \end{array} \right\}, \{< \mathbf{sc_m}, \mathbf{sc_j} >\} \right) \\ \hline \end{array} \right\rangle
\end{array} \right\}
$$

$$
\text{b.}\ \left\{
\begin{array}{l}
\left\langle [ENT, sc_m], \begin{array}{|c|}\hline \\ \hline \text{person}(sc_m) \\ \text{Named}(sc_m{,}Santa) \\ \hline \end{array}, \mathcal{K}_{sc,m} \cup \left\{ \begin{array}{|c|}\hline e_{utt,j} \\ \hline e_{utt,j} \prec n \\ e_{utt,j}\text{: men}(j_m{,}S'a{,}sc_m) \\ \hline \end{array} \right\}, -real \right\rangle \\[3em]
\left\langle [ENT, j_m], \begin{array}{|c|}\hline \\ \hline \text{person}(j_m) \\ \text{Named}(j_m{,}Johnny) \\ \hline \end{array}, \mathcal{K}_{m,j}, +real \right\rangle \\[3em]
\left\langle BEL, \begin{array}{|c|}\hline s_{j,bel} \\ \hline n \subseteq s_{j,bel} \\
s_{j,bel}{:}Att \left(j_m, \left\{ \begin{array}{l}
< [ENT, sc_j]\ , \\[1em]
\begin{array}{|c|}\hline \\ \hline \text{person}(sc_j) \\ \text{Named}(sc_j{,}Santa) \\ \hline \end{array}\ , \\[2em]
\mathcal{K}_{sc,j} \cup \left\{ \begin{array}{|c|}\hline e_{utt,m} \\ \hline e_{utt,m} \prec n \\ e_{utt,m}{:}\mathbf{men}(m_j{,}S'a{,}sc_j) \\ \hline \end{array} \right\}, \\[2em]
\hfill +real >
\end{array} \right\}, \left\{ <\widehat{\mathbf{sc}}_{\mathbf{j}}, \widehat{\mathbf{sc}}_{\mathbf{m}}> \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

In both (3.11.a) and (3.11.b) a belief has been added to the MSDs in (3.10.a) and (3.10.b). An almost all respects these beliefs are mirror images of each other. But there are two exceptions to this (marked in boldface in (3.11.a) and (3.11.b)). Mummy's belief in (3.11.b) attributes to Johnny an Entity representation for Santa that he has updated with a vicarious anchor in response to her use of *Santa*. But as recorded by the single member of the link set in 4th argument position of the *Att*-predication that makes up the bulk of the content specification of her belief, she interprets this update as only establishing a D-Link between Johnny's ER for Santa and her own – naturally, since she knows that Santa isn't a real person (which is reflected by the -real status of her *Santa*-labeled ER), so there can be no question of coreference between that ER and Johnny ER for Santa.

Johnny's belief in (3.11.a) attributes to mummy an ER for Santa that has been updated with a vicarious anchor in response to his use of *Santa*, and that this has established – or, more likely, reinforced – the coreference between her Santa representation and his. He is wrong on both counts, of course, as he is in his attribution to mummy of a *Santa*-labeled ER that is +real, like his own. Mummy for her part, also attributes a +real ER for Santa to Johnny. But she is right in that, while knowing that he is wrong.

A different but related situation is that where mummy and daddy are talking about Johnny. They share with other adults the knowledge that Santa doesn't exist and also both know that Johnny thinks Santa does exist. Suppose that mummy says to daddy:

(3.12) Johnny hopes that Santa will bring him a toy polar bear.

How does daddy interpret this communication against the background assumptions that he and mummy share, so that he can accept this as a true statement about what Johnny hopes even though the is a sense in which he must consider this hope incoherent insofar as it is about something that it falsely presupposes to exist? Intuitively there may not seem to be much of a riddle here. But it is something for which MSDRT ought to have a correct analysis.

In principle we could offer this treatment right here. But it is more or less isomorphic to one that we will come to in the next section, where we will look at some examples involving the notorious non-denoting name *Vulcan*, for the planet that was supposed to explain some features in the motions of Mercury. So we will let the matter rest until then.

3.2 Back to the use of names in attitude reports

Our only example so far of proper names occurring in the complements of attitudinal verbs was (2.6.e) ('John believes that Mary is in Paris'). The account we have given of the semantics of this sentence is, we believe, the intuitively right one for this example (at least in the absence of some very special context about John's understanding of the names *Mary* and *Paris*). As far as we can see, this treatment of the semantic contributions made by names in attitude content specifications is the general norm: it is common for our thoughts to be about particular people and things and for others to attribute such thoughts to us and to use names of those people and things in their contrnt specifications of those thoughts. And this way of interpreting definite noun phrases in the complements of attitude verbs is not restricted to proper names; pronouns, demonstrative phrase and definite descriptions in these positions often require such interpretations too.

But if this is how nearly all occurrences of names and other referring noun phrases in the content specifications of attitude reports are interpreted, it isn't so always. In particular, it isn't so in many of the examples in the philosophy of language literature that have kept the community busy for decades, and in some cases for more than a century.

Before we look at some of those examples, first a general point that in our view has been far too often overlooked or not been given the attention that it deserves. As MSDRT has it, to understand what happens when names are used in communication between a source S and a recipient H involves both S and H: the resources that S has and that enables her to use the name and resources that H has to interpret it in accordance with S's intentions. But when a name is used in the specification of an attitude content attributed to some third agent C, then that adds C as a third player. Earlier direct and indirect interactions between these three players may have their effects on the roles names play in the production and interpretation of an attitude attribution that S makes to C for the benefit of H. Unraveling these effects isn't always easy. In this section and in Section 6 we look at what such effects can be like.

3.2.1 The case of the failing name: *Vulcan*

We start with an example, inspired by what Kripke has had to say about the name *Vulcan.*

(3.13) Le Verrier assumed that Vulcan was closer to the sun than Mercury.

There is a sense in which this sentence is true, one in which its opposite – 'Le Verrier assumed that Vulcan was farther from the sun than Mercury' – is not. If someone had asked Le Verrier at the time when he had postulated a planet he called *Vulcan* to explain certain deviations of the movements of the planet Mercury from its predicted course whether he assumed that Vulcan was nearer to the sun than Mercury, he would have answered affirmatively.[8]

[8]How true the story is about Le Verrier and Vulcan on which philosophers base their judgment that (3.13) is true seems open to some questions. We proceed on the assumption that the facts are as they are supposed to be in the familiar philosophical discussions, setting aside questions of historical accuracy.

But how can we account for the claim that (3.13) is true within the semantics we have been developing for attitude attributions?

The central part of an account why (3.13) is true is that Le Verrier's mental state at the time to which the given utterance of (3.13) refers contained (i) a *Vulcan*-labeled entity representation for a planet, and (ii) a belief that this planet is closer to the sun than Mercury. Initially this entity representation will have had a single anchor, of a type that we haven't yet encountered. It anchors the entity representation via the definition that Le Verrier himself had used to introduce the name. Others, who took the concept and name from Le Verrier, will have had *Vulcan*-labeled entity representations with vicarious anchors connecting them to Le Verrier's. And through communication with them Le Verrier's own entity representation can be expected to have also acquired vicarious anchors to the original definition-based one. In this last respect the story of *Vulcan* among Le Verrier and his early fellow travelers is much like what we have already observed in connection with N-labeled ERs where$N$ originated in some other way than through definition.[9]

We assume that after Le Verrier came up with his Vulcan hypothesis and before it was shown that there could not be a planet with the properties that Le Verrier had ascribed to Vulcan, the name *Vulcan* was used by others who shared Le Verrier's assumption that there was such a planet and that *Vulcan* was its name. In our first scenario for an utterance of (3.13) we assume that

[9] A definition that introduces a proper name N always takes the form of specifying that N be the name of an entity that is being specified as the unique satisfier of some definite description. (So the definition consists of (i) the specifying description and (ii) the name.) Such name introductions are not foolproof, since it is possible that the description doesn't have a unique satisfier. In such cases the name doesn't have.a referent, although the one who introduced it will normally assume it does, and so may many others. This is what happened in the case of *Vulcan*.
In science the practice of introducing names by definition is much more common than it is in daily life. (Although it isn't unheard of there either. For instance, parents can decide to give their next child the first name *Kim* even before it has been conceived. When conception then occurs, the introduction has been successful. An explorer may call the next island he will discover *Prince Edward Island*, which and where it may be. But these are somewhat marginal cases. In mathematics on the other hand, definitions are our only means for introducing proper names – e.g. lower case *omega* as the first infinite ordinal or π as the quotient of the circumference of any circle and its diameter. More common, both in science and outside it, than the introduction by definition of proper names is the introduction of 'common names' by this method. Such definitions involve plural descriptions or 1-place predicates instead of singular definite descriptions, and introduce nouns as a 'common names' whose extensions consist of all entities that satisfy the 1-place predicate or the plural description.

the speaker S and the addressee H both belonged to this group of 'Le Verrier-followers'. Like Le Verrier himself, S and H both have *Vulcan*-labeled entity representations with the feature +real and a non-empty anchor set consisting of vicarious anchors as well as, perhaps the definition-based anchor that is distinctive of Le Verrier's own *Vulcan*-labeled entity representation. Like Le Verrier's, both S's and H's *Vulcan*-labeled entity representations are defective: they do not properly represent.

Because S's *Vulcan*-labeled entity representation is defective, her use of *Vulcan* in (3.13) is defective too. She thinks that by using *Vulcan* she refers to its bearer. Furthermore, she also assumes that while the attributee Le Verrier also has a properly referring *Vulcan*-labeled entity representation, which is also incorrect. But that doesn't affect the correctness of the attribution S makes or the interpretation that H constructs from her utterance. In fact, one might say that S and H are in an optimal position to make correct attributions to Le Verrier in relation to *Vulcan*, since the three of them are perfectly in tune with each other. In MSDRT terms, this means that S will represent the belief about Vulcan she attributes to Le Verrier in the same way as she would represent it as her own belief. That is, in both cases the representation of the belief itself will be one item in a larger MSD which also contains ERs for the entities that the belief is about. Furthermore, H's interpretation of (3.13)) can also be expected to take this form. This MSD is shown in (3.14).

$$
(3.14)\quad \left\{ \begin{array}{l}
\left\langle [ENT, v_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_L) \\ \text{Named}(v_L, \textit{Vulcan}) \\ \hline \end{array}, \left\{ \begin{array}{|c|}\hline \\ \hline \text{DEF: } \delta_{Vulcan}) \\ \hline \end{array} \right\}, +real \right\rangle \\[2em]
\left\langle [ENT, m_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\[2em]
\left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{star}(s_L) \\ \text{Named}(s_L, \textit{the Sun}) \\ \hline \end{array}, \mathcal{K}_{s,L}, +real \right\rangle \\[2em]
\left\langle BEL, \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: Closer-to'}(v_L, s_L, m_L) \\ \hline \end{array} \right\rangle
\end{array} \right\}
$$

This MSD is much like both the ‘inner’ MSD and the ‘outer’ MSD in (2.53), the representation that results in the mind of the recipient H from his interpretation of the attitude attribution “John believes that Mary is in Paris”. (The ‘outer’ MSD is (2.53) as a whole, the ‘inner’ MSD the argument of *Att*.) The similarity is made more explicit in (3.15), the MSD for the relevant part of H’s mind resulting from his interpretation of S’s utterance of (3.13).

$$
(3.15)\quad \left\{ \begin{array}{l}
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{speaker}(s_H) \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, l_H], \begin{array}{|c|}\hline \\ \hline \text{astronomer}(l_H) \\ \text{Named}(l_H, \textit{Le Verrier}) \\ \hline \end{array}, \mathcal{K}_{l_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{sun}(l_L H) \\ \text{Named}(l_H, \textit{the Sun}) \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\[2ex]
\left\langle [ENT, v_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_H) \\ \text{Named}(v_H, \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_H}, +real \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline t = n \quad t \subseteq s \\
s\colon\ Att \left(l_H, \left\{ \begin{array}{l}
\left\langle [ENT, m_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\
\left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_L) \\ \text{Named}(s_L, \textit{the Sun}) \\ \hline \end{array}, \mathcal{K}_{s,L}, +real \right\rangle \\
\left\langle [ENT, v_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_L) \\ \text{Named}(v_L, \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_L}, +real \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\colon\ \text{Closer-to'}(v_L, s_L, m_L) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{c} < s_L, s_H > \\ < m_L, m_H > \\ < v_L, v_H > \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array} \right\}
$$

The difference between (3.15) and (2.53) is that in the scenario assumed in connection with (2.53) all ERs were justified: H's own ERs for John, Mary and Paris properly represent these entities, and his representation of John's mental state is correct in two respects: (i) John does have entity representations for Mary and Paris, and (ii) these entity representations also properly represent (viz. the entities Mary and Paris). (3.15) differs in that this is not true for the *Vulcan*-labeled ERs. Neither of these has a referent, the recipient H's own ER no more than the ER that is part of his attribution to John.

And furthermore, in the scenarios considered all of this applies to S just as it applies to H: (3.15) and (2.53) also correctly represent the states of mind of the speaker, which are the basis for the attributions made, to John and to Le Verrier.

Note well that the misconception that S and H share with Le Verrier in the case we are considering does not affect – and more specifically, it doesn't negatively affect – the correctness of the attribution to Le Verrier that is made by (3.13). Whether the attribution is correct has to do with something that has not yet made an appearance in our formalization of these cases, viz. the mental state the attributee is actually in. The correctness of the attributions, as represented by the MSDs occurring as *Att*-arguments, is a matter of how those MSDs are related to those actual mental states of the attributees. Our formalization hasn't yet reached the point where we can speak of those actual states with the formal precision needed to say more about the correctness of attributions. That requires the model theory for MDRT that is developed in Chapter 4. It will be only in Section 6 of that chapter that we will return to the correctness of attitude attributions that involve non-representing ERs and vacuous names.

In the scenario for the use of (3.13) so far considered, S and H are like Le Verrier in that all three are under the illusion that Vulcan exists. But suppose now that S and H are not subject to this misapprehension – they are both aware that Le Verrier was betting on the wrong horse and that there is no such planet as Vulcan. We now assume that S's saying (3.13) to H takes place much later, after the non-existence of Vulcan has been demonstrated and universally accepted. This second scenario is similar to the one we mentioned at the end of Section 1.1, with mummy, daddy and Johnny. There is a kind of isomorphism between the two scenarios, with S corresponding to mummy, H to daddy and Le Verrier to Johnny; S and H share a +real ER for Le Verrier and a -real ER for Vulcan. Furthermore they share the belief about Le Verrier that he has a +real ER for Vulcan, and their own *Vulcan* ERs they take to be D-Linked to that ER. That is, before S makes her utterance of (3.13), the relevant part of H's mental state is given by the MSD shown in (3.16).

(3.16)

$$
\left\{
\begin{array}{l}
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{speaker}(s_H) \\ \hline\end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, l_H], \begin{array}{|c|}\hline \\ \hline \text{astronomer}(l_H) \\ \text{Named}(l_H,\ \textit{(Le Verrier)}) \\ \hline\end{array}, \mathcal{K}_{l_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{sun}(l_L H) \\ \text{Named}(l_H, \textit{(the Sun)}) \\ \hline\end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline\end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\[2ex]
\left\langle [ENT, v_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_H) \\ \text{Named}(v_H, \textit{Vulcan}) \\ \hline\end{array}, \mathcal{K}_{v_H}, -\mathbf{real} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline t = n \quad t \subseteq s \\
s\!: Att \left(l_H, \left\{
\begin{array}{l}
\left\langle [ENT, m_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline\end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\
\left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_L) \\ \text{Named}(s_L, \textit{the Sun}) \\ \hline\end{array}, \mathcal{K}_{s,L}, +real \right\rangle \\
\left\langle [ENT, v_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_L) \\ \text{Named}(v_L,\ \textit{Vulcan}) \\ \hline\end{array}, \mathcal{K}_{v_L}, +real \right\rangle
\end{array}
\right\} \right) \\ \hline\end{array} \right\rangle
\end{array}
\right\}
$$

It is against this background that H must interpret S's words. Let us assume that H interprets (3.13) as a belief attribution to Le Verrier and that he takes S's word for it that what she says is true. This means that the content of his interpretation will be a belief that he attributes to Le Verrier and that, since he accepts S's statement as true, this belief attribution to Le Verrier is adopted by H as a new belief. One way to represent this is to extend the mental state attribution to Le Verrier that forms the content of H's belief in (3.16) (and that consists of the three ERs for Mercury, the Sun and Vulcan) with his (H's) representation of S's attribution. In displaying the result in (3.17) we only show the updated belief of H (the update, that is, of the last member of the MSD in (3.16)).

(3.17)

$$\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \\ \hline t = n \quad t \subseteq s \\ s\text{: } Att \; \left(l_H, \left\{ \begin{array}{l} \left\langle [ENT, m_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\ \\ \left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_L) \\ \text{Named}(s_L, \textit{the Sun}) \\ \\ \hline \end{array}, \mathcal{K}_{s,L}, +real \right\rangle \\ \\ \left\langle [ENT, v_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_L) \\ \text{Named}(v_L, \textit{ Vulcan}) \\ \\ \hline \end{array}, \mathcal{K}_{v_L}, +real \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: Closer-to'}(v_L, s_L, m_L) \\ \\ \hline \end{array} \right\rangle \end{array} \right\}, \left\{ \begin{array}{c} < s_L, s_H > \\ < m_L, m_H > \\ < \hat{\mathbf{v}}_{\mathbf{L}}, \hat{\mathbf{v}}_{\mathbf{H}} > \end{array} \right\} \right) \\ \\ \hline \end{array} \right\rangle$$

(3.17) makes good sense, we believe, of the intuition that S's statement is true and that H is right to accept that it is, extending his belief state accordingly. The representation in (3.17) of the content of the belief attributed to Le Verrier is, one would expect, pretty much how Le Verrier would have represented it (or how he in fact did): as a belief about a planet he had named Vulcan. But at the same time it is clear from (3.17) that H is aware that strictly speaking there cannot be such a belief, since it is supposed to be an existing planet, but no such planet does in fact exist. The full MSD of which (3.17) is a member (cf. (3.16)) makes this clear through the D-Link between the *Vulcan*-labeled ER that H attributes to Le Verrier and the -real *Vulcan*-labeled ER that he has himself.

There is a little more to say about this second scenario. In our discussion of what Johnny says to mummy in Section 1.1 of this chapter and what mummy does with this utterance we assumed that she adds a D-Link to her -real ER for Santa. (The Link could only be a D-Link: you cannot add a vicarious anchor to the third component of a ?real ER, for that presupposes that the ER properly represents and that is precisely what the feature '-real' denies.) Likewise we can expect daddy to add a D-Link to his -real ER for Santa in his processing of mummy's utterance of (3.12). And once more by the. same token one should expect the H of our second scenario to D-Link his -real ER for Vulcan with the -real ER for Vulcan that he assumes S to have. This change in H's mental state isn't visible in (3.17) because H's own ER for Vulcan isn't shown there. But we hope that after all that has been said it is clear what we assume will happen to this ER: its third component gets as additional element the DRS $\boxed{\begin{array}{c} e' \\ \hline e' \prec n \\ e'\text{: men}(s_H, Vulcan, v_H) \end{array}}$.

If the scenario is as we have described it, this D-Link arguably makes not much of a difference to the account we have given. H makes use of his ER for Vulcan to represent the content of S's attribution to Le Verrier, and when the relevant part of his mental state is as described in (3.16), then he will take the attribution to be about the (presumed) referent of what he takes to be Le Verrier's ER for Vulcan, as shown in (3.16), with its D-Link to H's own -real Vulcan-ER.[10] But in our description of the case, in which the relevant

[10]How H arrives at this choice for his interpretation of S's attribution to Le Verrier is

background of H includes all that (3.16) displays, we have made somewhat easier for ourselves than they could have been. Suppose for instance instead that all H knows in connection with the name *Vulcan* is that it is the name that someone once gave to a planet, but that it then turned out there wasn't such a planet and so the name doesn't refer properly (though perhaps it is a bit of a mystery why anybody would still remember this). So H would have, we assume, a -real ER for a planet, but that is it. And then S comes along and says (3.13) to him. What is he to make of that?

Here is a sketch of a reconstruction. H can be expected to assume that S too has a a -real *Vulcan*-labeled ER. (Her use of the name *Vulcan* indicates that she has a *Vulcan*-labeled ER of which she relies in her use and it seems reasonable for H in the light of the little he knows that this ER must be, like his own, -real. Then the next clue for him is that the occurrence of *Vulcan* in (3.13) is inside the complement clause of the attribution it makes. If, as we have assumed, nothing more is known to H at this point, and he has never heard about any one called Le Verrier, then he will have to accommodate S's use of this name too. There is nothing in what we have said so far that predicts whether he will assume that this is the name of a real person and that S uses it as such in (3.13, i.e. that she has a +real *Le Verrier*-labeled ER that she has used in her use of *Le Verrier*, or whether S has used this name too as the name of a fictitious person. In fact, it doesn't matter much for the remainder of the story which of these two options H goes for, but let us assume that he takes *Le Verrier* to be the name of a real person and that he accommodates a +real *Le Verrier*-labeled ER with a vicarious anchor to the +real ER that he assumes has been used by S.

What next? When we first discussed occurrences of proper names in the complement clauses of attitude attributions in Section 2.4 of Ch. 2 we observed that such uses do not entail that the attributee must have an Entity Representation for the referent of the name. The only implication, we argued, was that she or he must have an ER for that referent and that the attitude attributed to her or him must have had a propositional content that is singular with respect to the entity represented by that ER. But is this an option when the speaker's ER can only be D-Linked to the ER of the attributee (as must be the case, since S's *Vulcan*-labeled ER is -real)? We aren't quite sure of this, and that is because we haven't explored as deeply as is

something that hasn't been spelled out in final detail here either. Let as assume that the MSD in (3.16) is exhaustive with respect to *Vulcan*-labeled ERs: H has only the one such ER that is shown in (3.16).

necessary for an answer how D-Links between +real and -real ERs of different agents can come about. As things stand we are inclined to think that it is very hard for ERs without proper referents to get D-Linked without the intermediacy of names than it is in the case of ERs that have proper referents.[11]

But to continue the story about H: Though we are unsure what guides interpreters in situation of the envisaged sort, one option is to assume that the attribution that S was making to to Le Verrier with her utterance of (3.13 is of the belief that a planet represented by the attributee Le Verrier in the form of a *Vulcan*-labeled ER is closer to the sun than Mercury, that this ER must have been +real – for real person can entertain such a belief about something that doesn't exist and that S must be thinking that this belief is in fact incoherent since her own ER for Vulcan is, like H's, -real.

There are quite a few iffy pints in this story. And even if the story is accepted as we have told it, the MSDRT representations we have thus far been using may not be expressive enough for a complete formalization. But this is as far as we can reach out to Vulcan here and now. In the next section we return to an example that involves properly representing ERs and vicarious anchoring, demonstrating more clearly than that was possible in Ch. 2, Section 2.4 that vicarious anchoring is perfectly possible in the absence of proper names.

3.2.2 Vicarious Anchoring without Proper Names

In the last part of the last section we dwelt on the problem of D-Linking between non-referring ERs in the absence of names or definite noun phrases that can take on the part of a proper name. When ERs do properly represent, however, and can be linked via vicarious anchoring, the presence of a name as linking vehicle, though common, isn't essential. As recalled in the final sentence of the last section, the observation was made in our discussion of

[11]This is a little too strong perhaps. The D-Linking we observed in our discussion of the case of the house on the Cornish coast in Section 1 of the present chapter was driven by the fact that the agents involved – the woman who had finally succeeded to build the house and the addressee whom she hadn't seen since before the building got under way – could recognize the individual concept ERs that their conversation partners were relying on because they recognized the shared description used (*that house we wanted to build on the Cornish coast*). In such cases the descriptions acquire the status of labels of the individual concept ERs that contain them as all or part of their descriptive material, as quasi- or ad hoc names of sorts.The question how agents share and communicate about individual concepts is a matter that needs closer investigation generally. The matter is especially important in connection with joint deliberation and planning, and will be taken up in the final chapter of the book.

"John believes that Mary is in Paris" in Section 2.4 of Ch. 2 that the implied ER in the mind of the attributee need not be labeled by the name used by the speaker. No decisive example of this possibility, though, was given there. the point is cliched by examples like (3.18.b), to which we turn here.

(3.18)

a. One of the people at work, who saw you at the party last night, came all the way to the seventh floor today to ask if I knew who you were. He seems to think you are pretty cute. But I said I couldn't give him your name or address without asking you first.

b. One of the people at work, who saw Ellen at the party last night, came all the way to the seventh floor today to ask if I knew who she was. He seems to think she is pretty cute. But I said I couldn't give him her name or address without asking her first.[12]

In both (3.18.a) and (3.18.b) it is plain from the first sentence that the man S is referring to has an entity representation for the person that has drawn his attention at the party, but one that is lacking a proper name as label. But that doesn't prevent S from making the attribution to him that she wants to make. In (3.18.b) S can do this by using the proper name *Ella* to refer to this person, which shows that there is no need for the entity representation she assumes her attributee (the man) to have for Ella to share the name *Ella* with her own entity representation for Ella, which is *Ella*-labeled. This is proof that the Reference-Based Principle does not require that the attributee's entity representation that is presupposed by the speaker's use of a name N must be labeled with N.

(3.18.a) is not directly relevant to the issue that (3.18.b) settles, but we have thrown it in since it is a particularly natural illustration of the possibility for an attributor S to refer to an entity that she assumes her attributee to have an entity representation for by using a noun phrase that makes no claim about the form of the attributee's entity representation; all that matters is the entity it represents. S's use of the pronoun *you* in (3.18.a) is clearly just for the benefit of the addressee – to enable her to understand that she is the object of the man's interests. But the example is not directly relevant to our discussion insofar as *you* isn't a name at all. It could not serve as a label for

[12]Think of (3.18.b) as said by S to H as a bit of gossip: S is speaking to H about their joint friend Ella, for which they both have entity representations, which are presumably vicariously linked with each other many times over.

any entity representation.

We won't make the considerable effort here of constructing DRSs that result from the interpretation of the full discourses (3.18.a) and (3.18.b). But there is one aspect of the interpretation of (3.18.a) that we want to say something about. It concerns the first two sentences of (3.18.a). The addressee H will interpret these sentences as saying something about her, because of the presence in them of the pronoun *you*. Assuming once again that H believes what S is saying to her, the relevant part of her mental state after interpretation of the two sentences might be as represented in (3.19).

(3.19)

$$\left\{ \begin{array}{l} \left\langle [ENT, sp], \begin{array}{|c|}\hline \\ \hline \text{woman}(sp) \\ \hline \end{array}, \mathcal{K}_{sp} \cup \left\{ \begin{array}{|c|}\hline s \\ \hline n \subseteq s \\ s\text{: talk-to}(sp.i) \\ \hline \end{array} \right\}, +real \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|} \hline t \quad e \quad t' \quad s' \quad m \\ \hline t \prec n \quad e \subseteq t \quad \text{man}(m) \quad e\text{: come-to-see'}(m,sp) \\ t' = n \quad t \subseteq s' \\ s'\text{: } Att \left(m, \left\{ \begin{array}{l} \left\langle [ENT, g], \begin{array}{|c|}\hline \\ \hline \text{woman}(g) \\ \hline \end{array}, \mathcal{K}_g, +real \right\rangle \\ \left\langle JUDGE, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: cute'}(g) \\ \hline \end{array} \right\rangle \end{array} \right\}, \{ <g,i> \} \right) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

(We haven't made much of an effort in (3.19) to represent the details of the first sentence, whose content has been reduced to there having been an event of some man coming to the speaker. Our central concern is with the second sentence.)

As far as the interpretation of this second sentence is concerned, the interpretation shown in (3.19) is one of two possibilities. (3.19) is the interpretation that would result when H has no recollection of the man that S is referring to.[13] He apparently saw her, but she doesn't remember anyone that might be him. But for all that (3.18.a) tells us, it is also possible that H did notice the man. Perhaps they exchanged a few words; perhaps she noticed his stare. In that case she will presumably still have an entity representation for the man and may recognize S's use of *one of the people at work* as presumably referring to the referent of that entity representation. In that case the mental state resulting from her interpretation of the first two sentences would be as in (3.20).

(3.20)

$$\left\{ \begin{array}{l}
\left\langle [ENT, sp], \begin{array}{|c|}\hline \\ \hline \text{woman}(sp) \\ \hline \end{array}, \mathcal{K}_{sp} \cup \left\{ \begin{array}{|c|}\hline s \\ \hline n \subseteq s \\ s\text{: talk-to}(sp,i) \\ \hline \end{array} \right\}, +real \right\rangle \\[2ex]
\left\langle [ENT, m], \begin{array}{|c|}\hline \\ \hline \text{man}(m) \\ \hline \end{array}, \mathcal{K}_{m}, +real \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline t \quad e \quad t' \quad s' \\ \hline t \prec n \quad e \subseteq t \quad e\text{: come-to-see'}(m,sp) \\ t' = n \quad t \subseteq s' \\ s'\text{: } Att \left(m, \left\{ \begin{array}{l} \left\langle [ENT, g], \begin{array}{|c|}\hline \\ \hline \text{woman}(g) \\ \hline \end{array}, \mathcal{K}_{g}, +real \right\rangle \\ \left\langle JUD, \begin{array}{|c|}\hline s' \\ \hline n \subseteq s' \\ s'\text{: cute'}(g) \\ \hline \end{array} \right\rangle \end{array} \right\}, \{ <g,i> \} \right) \\ \hline \end{array} \right\rangle
\end{array} \right\}$$

An additional reason for going into these details about the possible interpretations of (3.18.a), although it doesn't tell us anything about the use of names in attitude content specifications, is that it is a further example of

[13]That indefinite NPs, like *one of the people at work*, can be used to refer is something that linguists have been aware of for a long time. (See in particular (Farkas 2002). For a treatment of specific indefinites within MSDRT see (Kamp & Bende-Farkas 2019).

another topic of these notes: the benefits of the communication-theoretic approach to semantics for which MSDRT provides the foundations: Whether H construes the content of what S is telling her as an existential proposition, as in (3.19), or as a singular proposition about a man known to her, as in (3.20), depends on what information is available to H that she can bring to bear on her interpretation of S's utterance.

Exercise

Construct representations for the possible interpretations of (3.18.b).

Extend the representations in (3.19) and (3.20) with H's interpretation of the third sentence of (3.18.a). Likewise for your own construction of H's interpretation of the first two sentences of (3.18.b).

After this interlude about the Reference-Based Principle we briefly return to the empty name problem exemplified by *Vulcan*, but only to introduce a name for the principle that justifies the use of empty names like *Vulcan* in attitude attributions. In analogy with the label 'Reference-Based Principle' we call the principle that justifies S's use of *Vulcan* in (3.13) the 'Link-Based Principle'. And we repeat the open question connected with this principle: The name uses justified by the principle that we have come upon so far all have the following feature: the entity representation that is presupposed to be part of the mental state of the attributee by the speaker's use of the name has that namer as a label. But is that a necessary constraint on applications of the Link-Based Principle? Here is an attempt at an answer. It starts with a comment on the representations in (3.19) and (3.20).

As given, (3.19) says nothing about the anchor set of the man's entity representation for H; it just uses the dummy symbol '$\mathcal{K}_g$' to indicate that the ER does have one or more anchors. However, in the case at hand the addressee will be in a position to make more specific assumptions about this anchor set. From the first sentence of (3.18.a) she has learned that there will be at least one perceptual anchor in this set, a witness to the man noticing her at the party. This perceptual anchor (or several, if he noticed her repeatedly) is different from the perceptual anchors that were introduced in Chapter 2. Those were witnesses to perceptions *as* those perceptions were going on. In contrast, the perceptual anchor or anchors that the addressee can infer to be part of the man's entity representation of her are witnesses of one or more *past* perceptions.

Moreover, the addressee may feel entitled in this case to assume that this anchor is one that the man continues to be fully aware of. But is being consciously remembered a general property of anchors? Our view is that it isn't. But the question to what extent consciousness is required for the presence of anchors is a non-trivial one, which needs careful discussion. This is not the right point to go into this, but the matter will have to be addressed at some point.

Past perceptual anchors also raise another issue: When are current perceptual anchors turned into past perceptual anchors? This is a special case of the more general question when things that we take to be the case currently are to be recast as things that obtained earlier, at the time when we became aware that they were going on, but that need not be the case any longer. Our cognitive systems have built-in strategies for making such transitions – from what we take to be the case currently to what we know was the case earlier but is perhaps no longer now – at some point after we had direct access to the relevant information (e.g. through perception). Part of those strategies is to make such transitions as a function of the information involved.[14]

To represent a past perceptual anchor for the man's entity representation of her, the addressee of (3.13) must assume that he has an entity representation for the party that is linked to her own entity representation for it and that the anchor of his entity representation for her locates the past perception of her at that party. If p_m is the distinguished dref for the entity representation of the party that she attributes to the man and g_m the distinguished dref for the entity representation of herself that she attributes to him, then the past perceptual anchor she attributes to him can be assumed to have the following form:

$$(3.21) \quad \begin{array}{|c|} \hline e \\ \hline e \prec n \quad \mathrm{AT}(e,p_m) \\ e\text{: see}(i,g_m) \\ \hline \end{array}$$

Here the predicate 'AT' expresses a relation between events which holds when the first event occurred 'at' the second, not only in the sense that the first is temporally included in the second, but also in the sense of spatial inclusion. Exercises 1. The full modification of (3.19) that is needed to implement this more explicit treatment of the anchor set of the man's entity representation

[14] Think of roughly for how long on after observation we hold on to it still being the case that (i) there is a bird on the roof; (ii) there is a cat on the roof; (iii) there is a dead branch on the roof.

of the addressee is left as an exercise.

A second exercise is to reconstruct H's interpretation of (3.18.a) on the assumption that she takes the indefinite *one of the people at work* as an epistemically specific indefinite. (This involves her accommodating an entity representation for the person that she takes S to be referring to and vicariously anchoring it to the entity representation that she takes S to have for him.)

And thirdly a further, more elaborate exercise: Give a systematic reconstruction of the interpretation of all of (3.18.a). This involves a number of problems that have not been discussed in this document. But give it a try. It helps to understand how the MSDRT framework works, while making one aware of how much is still missing.

[end exercises]

This is the end (for now) of our discussion of non-referring names in the complement clauses of attitude attributions. Our next examples are of a very different sort: They involve the case where the attributee has several names for the same entity without knowing that they do; or – the opposite end of the wrong stick – treat a name for two different entities as if it were a name for just one.

3.3 Misalignments between Names and their Referents

The next few examples are all instances of the same basic pattern: the attributee has two distinct entity representations for the same entity without knowing that they are representatives of the same entity, and in fact in most of the examples they are convinced that the represented entities are distinct. As time goes on, they may associate more information – a growing set of beliefs and perhaps other attitudes as well – with these respective entity representations and in the examples that philosophers have found the greatest challenges some of those beliefs may formally contradict each other. (Phosphorus can only be seen shortly before dawn, Hesperus can only be seen shortly after sunset; Londres is beautiful, London is ugly; Superman flies, Clark Kent doesn't; fewer people know about Tully than about Cicero.)

The problem with these examples is not that the entity representations involved fail to properly refer. They do refer properly (and so do the attitude content representations in which they participate via their distinguished discourse referents). The problem is that there are two entity representations where there really ought to be only one. That prevents the agent from making connections that he would have been able to make if it hadn't been for this and would nave protected him from assumptions about combinations of properties that could not all be simultaneously true of a single entity.[15]

There are two sides to the problems such multiple ERs for the same entity can create. In the mental state of the agent who has two entity representations for the same entity without realizing this, it creates the misrepresentation of reality described above. But there is also a problem for a speaker S who wants to describe the mental states of agents who suffer from this predicament. What are the forms she can use in attitude attributions to such agents that do justice to their predicament by conveying a correct picture of their skewed mental states. Among other things this requirement imposes constraints on the use of proper names in such attributions. It should be clear that these constraints also concern the attributor's interlocutor H. H must be in a position to interpret the names that S uses to point to one or the other of the two entity representations of the attributee that he falsely sees as representations of distinct entities. This difficulty manifests itself especially when S and H themselves have a single representation for the entity for which the attributee has two.

All this is well-trodden ground (so much so that sometimes one cannot help feeling that nothing worth having will ever grow on it again). The main moral to retain from this description of the problems is that a viable account of attitude attributions to agents suffering from double vision must deal with two issues:

(i) It must be based on a correct description of the mental states of such agents.

(ii) It must describe the constraints on the forms that attributors can use to express the attributions they want to make to such agents, so that their interlocutors can reconstruct the relevant parts of the agent's mental state from the attitude report and interpret the attributed attitudes accordingly.

The solution that MSDRT offers for (i) is obvious. But nothing we have said about MSDRT so far shows how to deal with (ii). This will be the focus of

[15]The predicament of such agents has been aptly described as 'Double Vision' (Klein 1981) [**Check Reference to Ewan Klein**].

the present section.

We start with reports about people who didn't know that Phosphorus (or 'the Morning Star') is the same celestial body as Hesperus (or 'the Evening Star'). Historically, the discussion of Hesperus-Phosphorus cases started with Frege's observation that although these names have the same referent ('Bedeutung'), they cannot be safely substituted for each other *salva veritate* in belief and other non-extensional contexts, a point often illustrated by sentences like (3.22.a,b).

(3.22)a. The ancient Egyptians[16] believed that Phosphorus is visible some time before dawn, but at no other time.

b. The ancient Egyptians believed that Hesperus is visible some time before dawn, but at no other time.

c. The ancient Egyptians believed that Venus is visible some time before dawn, but at no other time.

The point here is that the first sentence has a claim to being true that the second does not. Moreover, it isn't quite clear what one is to make of (3.22.c), in which *Venus* occupies the position of *Hesperus* and *Phosphorus* in (3.22.a) and (3.22.b). There is an intuition that (3.22.c) doesn't make a coherent claim at all.

What is the basis for these judgments? At the level of discussion adopted so far in this section the obvious thing to say would be this: We have learned, in high school perhaps or perhaps from reading Frege, that *Phosphorus* was used by the ancient Greeks as name for the celestial body that can sometimes be seen shortly before dawn and is then the brightest object in the sky and *Hesperus* for the celestial body that can sometimes be seen shortly after sunset and that is the brightest object in the sky then. For the Greeks *Phosphorus* and *Hesperus* were the names of distinct objects; and within philosophy these names have become conventionalized as devices for referring to the distinct conceptions they had of what we now know to be one and the same object. This convention is used also in (3.22), although those sentences

[16]We haven't said anything about plural noun phrases so far. This is a big topic of which we will discuss only those aspects that will prove necessary for what we want to say about joint deliberation and action. For some of those aspects see Ch. 6. At this point it doesn't matter too much how the predication 'the Egyptians believed that' is interpreted. It suffices for present purposes to think of the Egyptians as a composite individual with a single mental state.

speak not about Greeks but about Egyptians, who unsurprisingly used entirely different names: *Tioumoutiri* for *Phosphorus* and *Ouaiti* for *Hesperus*.

One of the tasks that we are presented with by our intuitions about what is right or wrong about the sentences in (3.22) is to explain the source of those intuitions and to determine the extent to which they can be justified. MSDRT is to help us with these tasks. But let us, instead of dealing directly with the sentences in (3.22), which present some complications that have to with the plural subject *the ancient Egyptians* and that are not germane to the problems we should be addressing here, consider variants involving a single Greek person of whom it is plausible to assume that he too thought that *Hesperus* and *Phosphorus* are two. We have chosen Thales, taking it as established that it was Pythagoras – younger than Thales by half a century – who first recognized that *Hesperus* was the same body as *Phosphorus*.[17] So instead of on (3.22) let us focus on (3.23). And let us also assume, although there may be less evidence for this assumption, that Thales used the names *Hesperus* and *Phosphorus* for what we take him to have taken to be two distinct bodies in the sky. Consider the sentences

(3.23)a. Thales believed that Phosphorus is visible some time before dawn, but at no other time.

b. Thales believed that Hesperus is visible some time before dawn, but at no other time.

c. Thales believed that Venus is visible some time before dawn, but at no other time.

To repeat what we already said about the sentences in (3.22): (3.23.a) appears to be true, (3.23.b) appears to be false; and it is not so clear what to make of (3.23.c).

The reconstruction in MSDRT of what goes on in exchanges involving sentences like those in (3.23) follows the same lines as we did in our treatment of the sentence about Vulcan in (3.13). Both S and H, we assume, have a picture of the relevant part of the mental state of Thales, according to which he had distinct entity representations which are linked to our uses of *Hesperus* and *Phosphorus*, with which he associated different beliefs and – this is

[17]It may also be taken for granted that Thales, who must have been an astronomer of considerable sophistication – he is generally credited with having predicted a solar eclipse in 585 BC – must have been aware of the existence of *Hesperus* and *Phosphorus* and would certainly have understood the question whether or not they were the same even if it never actually occurred to him.

a difference with the Vulcan case of the last section – which are anchored to the entity they represent by distinct perceptual anchors: anchors witnessing observations before dawn for his '*Phosphorus*'-labeled entity representation and anchors witnessing observations after dusk for his entity representation labeled with *Hesperus*. The attitudes attributed by such reports are then represented by both S and H as associated with the one or the other of those entity representations, depending on whether S uses *Phosphorus* or *Hesperus*. To repeat the point already made about double vision cases: Which of the two entity representations of Thales those attitudes are associated with is obviously essential to how he thinks about the entities represented by those two entity representations, and to the inferences about Venus he is able and willing to draw, and more generally to what speculations about Venus he is motivated to engage in. It is – using a term that we will make formally precise later – crucial for the *cognitive import* of an agent's representations, as going beyond the truth conditions that those representations determine. A good report of such thoughts should enable the recipient H to represent this cognitive import correctly. And the theory should be able to say which reports are good in this regard and which are not.

The use among philosophers of language, logic and mind of the names *Phosphorus* or *Hesperus* in phrasing reports like those in (3.22) and (3.23) is governed by convention – a natural convention arguably because of the obvious etymological connection between '*Phosphorus* and *the Morning Star* on the one hand and *Hesperus* and *the Evening Star* on the other; but a convention nevertheless. The extent to which this is so becomes clear when we try to use the name *Venus* instead, as in (3.23.c) and (3.22.c). We have no well-defined, generally accepted strategy for using and interpreting *Venus* in such sentences. Should (3.23.c) be taken as equivalent with (3.23.a) or as equivalent with (3.23.b)? Or as equivalent to the one in some contexts and to the other in some other contexts?[18]

[18] Here, for what it is worth, are our own intuitions about the sentences in (3.23.c) and (3.22.c). Because of no conventionalized bias towards one or the other of the two entity representations that Thales and others from his days had for what we now know are different manifestations of Venus, we take the name *Venus* as standing for our own concept, of Venus as a single entity, and this also when it occurs in sentences like (3.23.c) and (3.22.c). In view of that, (3.22.c) is false for us, and so is the sentence 'Thales believed that Venus is visible some time after dusk, but at no other time.' No better is 'Thales believed that Venus is visible some time before dawn and also some time after dusk,' although that is a belief that we ourselves subscribe to. True for us are the sentences 'Thales believed that Venus is visible some time before dawn' and 'Thales believed that Venus is visible some time after dusk.'

To sum up: The reason why communications like (3.23.a) and (3.23.b) work as well as they do depends on the high degree to which our use of the names *Phosphorus* or *Hesperus* is of conventionalized when we talk about those who thought the bearers of those names were distinct.

For an MSDRT treatment of the sentences in (3.23) everything is in essence already in place and we could have left these treatments as an exercise without a bad conscience. But there is a more general point that these treatments bring up and that hasn't been discussed yet. So we give a reasonably explicit treatment of those sentences, in order to have a concrete example of this more general observation.

As in our previous treatments of utterances with names, we adopt MSDRT's communication-theoretic approach in our treatment of these examples, by giving the relevant parts of the interpreter H's mental state before and after interpretation of the input utterance. But what in the cases before us should be considered the relevant part of H's state at the point when he gets the input sentence? This is the general issue that utterances of the sentences in (3.23) and (3.22) point us to. We already touched upon this when we observed that within certain groups of speakers today, to whom we count ourselves and also those we expect will have a look at this text, the use of *Phosphorus* or *Hesperus* is highly conventionalized. For those conventions to be operational the recipients of utterances like (3.23) and (3.22) must come to the process of interpretation with a certain dose of special preparatory information already in place – information that prepares the interpreter for these conventionalized uses of *Hesperus* and *Phosphorus*.

The form in which H's knowledge of the convention regarding *Phosphorus* or *Hesperus* could be represented is shown in (3.24).

(3.24)

$$\left\{ \begin{array}{l}
\left\langle [ENT,\alpha], \begin{array}{|c|}\hline \\ \hline \text{agent}(\alpha) \\ \hline \end{array}, \mathcal{K}_{\alpha_H}, +/-\mathit{real} \right\rangle \\
\left\langle [ENT,v], \begin{array}{|c|}\hline \\ \hline \text{planet}(v) \\ \text{Named}(v,\ \mathit{Venus}) \\ \\ \hline \end{array}, \mathcal{K}_{v_H}, +\mathit{real} \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline t = n \quad t \subseteq s \\
s\colon \mathit{Att} \left(\alpha, \left\{ \begin{array}{l}
\left\langle [ENT,\eta], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\eta) \\ \text{N'd}(\eta,\ \mathit{Hesperus*}) \\ \hline \end{array}, \mathcal{K}_{\eta_\alpha}, +\mathit{real} \right\rangle \\
\left\langle [ENT,\phi], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\phi) \\ \text{N'd}(\phi, \mathit{Phosphorus*}) \\ \hline \end{array}, \mathcal{K}_{\phi_\alpha}, +\mathit{real} \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \eta \neq \phi \\ \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{c} <\eta, v> \\ <\phi, v> \end{array} \right\} \right) \\ \hline \end{array} \right\rangle
\end{array} \right.$$

(3.24) is to be seen as a kind of schema, with α as schematic letter, which has to be instantiated when (3.24) is applied in interpretations of sentences like those in (3.22.a,b) and (3.23.a,b). In such interpretations the interpreter H will have to convince himself that the agent or agents to whom attributions are made fits or fit the schema in (3.24). But when he has convinced himself of this, then (3.24) will greatly facilitate his interpretation of the attribution that is being made. For instance, to interpret (3.23.a), H only needs to instantiate the scheme and then add to the MSD of the instantiated schemas his interpretation of the attributed belief.

The asterisks in *Hesperus** and *Phosphorus** in (3.24) stand for the names that α had as labels for his entity representations for the Morning Star and the Evening Star. The use of '*Hesperus**' and '*Phosphorus**' in (3.25) is meant to capture the conventionalization of *our* use of *Hesperus* and *Phosphorus* in talking about those who didn't realize they were one and the same. (3.25) gives the update of (3.24) after interpretation of (3.23.a). Representing the content of the belief attributed to Thales in (3.23.a) presents some problems, but ones have nothing to do with the use the name *Phosphorus*. In (3.25) these have simply been set aside, by adopting an ad hoc paraphrase. For further details see the comments below (3.25).

(3.25)

$$
\left\{
\begin{array}{l}
\left\langle [ENT,\theta], \begin{array}{|c|}\hline \\ \hline \text{agent}(\theta) \\ \text{N'd}(\theta,\ \textit{Thales}) \\ \hline \end{array}, \mathcal{K}_{\theta_H}, +real \right\rangle
\quad
\left\langle [ENT,v], \begin{array}{|c|}\hline \\ \hline \text{planet}(v) \\ \text{Named}(v,\ \textit{Venus}) \\ \hline \end{array}, \mathcal{K}_{v_H}, +real \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \\ \hline \\ t = n \quad t \subseteq s \\ \\
s\colon Att \left(\theta, \left\{ \begin{array}{l}
\left\langle [ENT,\eta], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\eta) \\ \text{N'd}(\eta,\ \textit{Hesperus*}) \\ \hline \end{array}, \mathcal{K}_{\eta_\theta}, +real \right\rangle \\
\left\langle [ENT,\phi], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\phi) \\ \text{N'd}(\phi,\textit{Phosphorus*}) \\ \hline \end{array}, \mathcal{K}_{\phi_\theta}, +real \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \eta \neq \phi \\ \hline \end{array} \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \text{sometimes-visible-up-to-} \\ \text{one-hour-before-dawn}(\phi) \\ \text{never-visible-after-dusk}(\phi) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{c} <\eta, v> \\ <\phi, v> \end{array} \right\} \right) \\ \\ \hline \end{array} \right\rangle
\end{array}
\right\}
$$

Legenda The representation in (3.25) of the content of the belief attributed to Thales by (3.23.a) has been suppressed and replaced by an ad hoc paraphrase of what is in the represented sentence. But a proper account of who this representation is derived would have required attending to various details about the processing of the complement clause of (3.23.a) and that would have led to a serious side track. For pointers to the literature on present-under-past sentences, of which (3.23.a) is an example, see Ch. 2, Section 2.

The most difficult problem from the perspective of explicit truth conditions is the inherent vagueness of propositional phrases like *after dusk* and *before dawn*. A first shot at the semantics of these phrases is this: for any calendar day d, the dawn and the dusk of d are certain periods within d, with dawn starting when it is getting light and ending (according to our understanding of the meaning of *dawn*) when the sun appears over the horizon, and with dusk starting when the sun disappears below the horizon and ending when all daylight has faded. But these notions are already affected by vagueness. There is however another and even more serious element of vagueness connected with the prepositional phrases *before dawn* and *after dusk*: In order for an eventuality to be describable as having occurred before dawn, how much time may have elapsed since its occurrence. Surely it shouldn't have been as much as ten hours, for then the event should have been described as *after dusk* (assuming dawn isn't over already). But then, how much less than ten hours? There is no simple answer to this question that we know of. [end Legenda]

We can sidestep these particular problems by assuming that dawn and dusk are particular sub-periods of calendar days, where the latter are 24 hour periods that, we will assume, last from midnight till midnight. That is, we introduce 'C.DAY' as a predicate of periods of time such that C.DAY(t) is true iff t is such a midnight till midnight stretch of time, and 'DAWN' and 'DUSK' as 2-place relations between periods of time where the second argument is always a calendar day and the first is that sub-period of it that qualifies as its dawn or dusk, respectively. And we just assume a well-defined limit on before and after by assuming that 'BEFORE' and 'AFTER' are also 2-place predicates of periods of time where 'BEFORE(t,t')' holds when t is a period that abuts t' on the left – i.e. $t \supset\subset t'$ – and is 'not too long', and 'AFTER(t,t')' means that t abuts t' on the right – $t' \supset\subset t$ and isn't 'too long'. Having thus set aside the real problems by brute force, one way in which we can approximate the attributed belief content by the following DRS is shown in (3.26).

$$
(3.26)\quad
\begin{array}{|c|}
\hline
d \quad t' \quad t'' \\
\hline
\text{C.DAY}(d) \quad \text{DAWN}(t',d) \quad \text{BEFORE}(t'',t') \\
\Diamond \begin{array}{|c|} \hline e' \quad p \\ \hline e' \subseteq t' \quad \text{person}(p) \\ e'\text{: see}(p,\phi) \\ \hline \end{array} \\
\neg \begin{array}{|c|} \hline d' \quad t^3 \quad t^4 \\ \hline \text{C.DAY}(d') \quad \text{DUSK}(t^3,d') \quad \text{AFTER}(t^4,t^3) \\ \Diamond \begin{array}{|c|} \hline e'' \quad p' \\ \hline e'' \subseteq t^4 \quad \text{person}(p') \\ e''\text{: see}(p',\phi) \\ \\ \hline \end{array} \\ \hline \end{array} \\
\hline
\end{array}
$$

Comments

1. Possibility is represented here by $\Diamond$, the standard device for this purpose in modal logic. We are not concerned here about the kind of possibility that is involved.

2. (3.26) says that there are days d when it is possible for a person to see ϕ at some time before dawn but no days at which it is possible to see ϕ after dusk. That representation misses the 'generic' flavor that some will connect with the sentences in (3.22) and (3.23): that it is generally possible to see Phosphorus before dawn when the days are right and not simply that there are some days when Phosphorus can be seen at that time of day. But arguably (3.26) isn't that far off. Some speakers are surprisingly naive about the elementary facts of basic astronomy. For them the representation in (3.26) may not be at variance with their understanding of the sentences in (3.22) and (3.23).

3.3.1 Using *Hesperus* and *Phosphorus* in unfamiliar attitude attributions

This much about representation of the sentences in (3.23). We conclude this discussion of Hesperus-Phosphorus sentences with an observation of another aspect that can be found in many of the discussions in the philosophical literature of examples like those in (3.23). Typically these discussions focus on sentences of which it is plain whether or not they are true. The sentences only serve to see if certain theories of the truth conditions of sentences of some given types agree with our intuitions. (3.23.a,b) – and up to a point also (3.23.c) – are examples of this. (In fact, we could plausibly have assumed about the pre-interpretation mental state of our interpreter H of (3.23.a) that it included the belief attributed to Thales from the start, in which case the attribution wouldn't have told the interpreter anything new.

It is worth noting, however, that the conventions we use in talking about ancient views about Hesperus and Phosphorus are also operative when what is said *is* meant to be genuinely informative. For instance, according to our sparse scattered layman's information about ancient Egypt, there were distinct 'Hesperus' and 'Phosphorus' cults, each presumably with their own caste of priests. Suppose that Chaem[19] is a priest from the reign of Ramesses II (ca. 1303 - 1213 BC). And suppose that some Egyptologist has just deciphered an old inscription from which he infers the claim in (3.27).

(3.27) Chaem was at first responsible for the Hesperus cult and only later for the Phosphorus cult.

This may be interesting news to Egyptologists working on the early part of the New Kingdom, and a trigger of further comments and disputes. What (3.27) says will not be a problem for those participating in the debate (though they would, as Egyptologists, look askance at the use of these particular names), nor what would be said if *Hesperus* and *Phosphorus* were exchanged (which would have been false on the assumption that (3.27) is true).

3.3.2 Kent and Superman

The degree to which we rely on the alignment of our use of names with that of those to whom we attribute certain attitudes is just as extensive when these attributees are fictional characters. This is so, for instance, in what

[19] Our apologies. Chaem is just made up by us. Expert Egyptologists would do better than we can.

may be the most beloved example of all in contemporary analytic philosophy in North America: the fraught relationship between Lois Lane and Clark Kent and his alias Superman. Those who actively participate in the philosophical debates surrounding this relationship all appear to be steeped in its paradoxical tragedies, to an extent that it is probably impossible to tell them anything that is true according to the story but that they do not know. Part of the story is, slightly adjusted in MSDRT terms, that Lois has *Superman*- and *Clark*-labeled entity representations that she takes to be the representations of different – in fact, VERY different – people. She keeps having new encounters with both these characters and each time she meets one of them, she adds a further anchor to the entity representation involved in her recognition of them and often also new attitudes associated with this entity representation – new beliefs, but, most of all, emotionally charged judgments and opinions.

Lois' (and Clark's) predicament is typically exploited for philosophical purposes by putting up for discussion sentences like (3.28).

(3.28)a. Lois thinks Superman is wonderful and she thinks Clark Kent is a fool.

b. Lois thinks Clark Kent is wonderful and she thinks Superman is a fool.

c. Lois thinks Superman is wonderful and she thinks Superman is a fool.

d. Lois thinks Clark Kent is wonderful and she thinks Clark Kent is a fool.

Superman initiates have very clear intuitions about the truth values of these attributions to Lois: (3.28.a) is true and the other three are false. Moreover, (3.28.c) and (3.28.d) have a flavor of absurdity about them. One feels that quite apart from being wrong attributions to Lois Lane, they are of a form that couldn't be a correct attribution to anyone.

No one more or less familiar with the Superman story would challenge these intuitions. But is there any way in which we can explain or theoretically support them? One way to proceed is to adopt the method followed so far in Section 3 of this chapter, that of considering particular utterance scenarios for the sentences and looking in each case how interpretation of the utterance changes the mental state of the interpreter.

But what are plausible utterance situations for these sentences? Here are two suggestions. The first is an imaginary utterance made by some inhabitant of Metropolis (the city where Lois Lane, Clark Kent and the other protagonists of the story live). Since more or less everybody in Metropolis knows Superman, while the double identity of Superman, as Superman and as Clark Kent, is known to him alone, we may assume that any person S who utters (3.28.a), with its occurrences of both *Superman* and *Clark Kent*, has a *Superman*-labeled entity representation and a *Clark Kent*-labeled entity representation both of which are anchored to Clark Kent/Superman. But S, like anyone else who knows Clark Kent as the hapless reporter of the Metropolitan Planet, will take the referents of these two entity representations to be different persons. Moreover, we may assume that S's *Superman*-labeled entity representation is directly or indirectly linked with the *Superman*-labeled entity representation of Lois, and that the same is true of S's and Lois' *Clark Kent*-labeled entity representations. And, equally important, S's *Superman*-labeled entity representation can be expected *not* to be linked to Lois' *Clark Kent*-labeled entity representation, and likewise for S's *Clark Kent*-labeled entity representation and Lois' *Superman*-labeled entity representation. Let us assume that these expectations are true. And let us also assume that Lois' mental state is as the story implies. (3.29) is an MSD for the relevant part of it.[20]

$$
(3.29)\quad \left\{ \begin{array}{l}
\left\langle [ENT, sm], \begin{array}{|c|}\hline \\ \hline \text{person}(sm) \\ \text{Named}(sm,\ \textit{(Superman)}) \\ \hline \end{array}, \mathcal{K}_{sm,S}, +real \right\rangle \\[2em]
\left\langle [ENT, ck], \begin{array}{|c|}\hline \\ \hline \text{person}(ck) \\ \text{Named}(ck,\ \textit{(Clark Kent)}) \\ \hline \end{array}, \mathcal{K}_{ck,S}, +real \right\rangle \\[2em]
\left\langle BEL, \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: wonderful'}(sm) \\ \hline \end{array} \right\rangle \quad \left\langle BEL, \begin{array}{|c|}\hline s_2 \\ \hline n \subseteq s_2 \\ s_2\text{: fool'}(ck) \\ \hline \end{array} \right\rangle
\end{array} \right\}
$$

Then, if S's uses of the names *Superman* and *Clark Kent* in the complement clauses of the occurrences of *think* are interpreted according to the Link-Based

[20]For DRS construction for sentences whose verb phrases are of the form 'copular *be* + adjectival or nominal complement' see (Kamp 2021*a*).

Principle (see Section 2.2), (3.28.a) will come out true, since its interpretation will be confirmed by the MSD in (3.29).
The case just considered is reminiscent of our first scenario for the *Vulcan* sentence (3.13) and the Phosphorus and Hesperus sentences in (3.23.a) and (3.23.b): S and H share the shortcomings of the mental state of the attributee. However, the cases differ in the nature of the shortcomings. In the case of Vulcan it is thinking one has an entity representation for something that proves to be illusory. In the case of Hesperus and Phosphorus and of Kent and Superman it is having too many representations for something that does exist. But otherwise the case before us is much like that where (3.13) is uttered in an exchange between contemporaries of Le Verrier's who share his erroneous belief that Vulcan exists. In particular, the difference between the real interlocutors S and H from our Vulcan scenario and the fictional interlocutors we are imagining right now. In both cases S's words and H's interpretation of them will proceed as if everything were as the mistaken belief that is shared between S, H and the attributee. In this scenario it may be assumed that S and H have *Superman*- and *Clark Kent*-labeled entity representations just like Lois and that their respective entity representations are vicariously linked, as well as coreferential. In this case the updates of H's mental state resulting from his interpretation of the sentences in (3.28) will be the result of H applying (among other rules) the Reference-Based Principle and will be as expected from our treatment earlier of the attribution sentence 'John believes that Mary is in Paris'. (For the truth-conditional import of that update we must refer once more to MSDRT's model theory, the presentation of which is still to come.)

Let us now turn to a scenario in which Lois' misconception of Clark Kent/ Superman is not shared by S and H. S and H are now real people, who know enough about the Superman story to be aware of Lois' predicament and in a position to use and understand the sentences in (3.28) the way they should. It seems a reasonable assumption that in this case too S's use and H's interpretation of the names *Clark Kent* and *Superman* in the sentences of (3.28) involve *Superman*- and *Clark Kent*-labeled entity representations and the vicarious links between those entity representations and the *Superman*- and *Clark Kent*-labeled entity representations in the mental state of Lois. But what exactly can vicarious linking mean between entity representations of real people like S and H and a fictional character like Lois Lane?

What answer can be given to this question depends very much on what one adopts as one's theory of fiction and fictional characters. In (Kamp 2021*b*) the proposal is made that those who are familiar with a piece of fiction, as the S and H of our present scenario for the use of sentences in (3.28) are assumed to be, will have entity representations for the protagonists of the fiction. In general, not all of those will be labeled with names, but of them many will be, and here we restrict attention to those. We assume that S and H both have a *Lois*-labeled entity representation for Lois Lane and a doubly labeled entity representation, with the labels *Clark* and *Superman*, for the person that they know to be one and the same.

N-labeled entity representations for protagonists from a given piece of fiction will be formed as part of 'partaking' in the story, by reading the text or watching the movie, and the N-labeled entity representations of those who get them this way are assumed to be linked: that is part of being joint participants in a piece of fiction. But such entity representation can be passed on to others who are not participants in this sense, by mentioning the name N in their presence, which will enable them to adopt the name, as a name for what it has just been used as the name of, in the manner described in our discussion of the *Gogol* example in Section 2.4 of Ch. 2. By these two mechanisms, for participants and for non-participants, networks of vicariously linked entity representations get established. (Kamp 2021*b*) proposes that these networks of vicariously linked entity representations can be identified with *fictional characters* and that, more particularly, the fictional character which goes by the name N is identified with the network that consists of all N-labeled entity representations. Thus fictional characters are abstract entities in our actual world, rather than concrete entities in the world of the fiction. Another consequence of this way of defining fictional characters and connecting them with the corresponding protagonists from the fiction is that the N-labeled entity representations of any two agents that have them will be 'vicariously linked' via the two mechanisms mentioned: the one that links the representations of participants and the mechanism involving vicarious anchoring that applies to participants and non-participants alike.

Those agents who have one of the N-labeled entity representations from the network may each attach various kinds of information with it (in the form of DRSs in which the distinguished dref of the entity representation figures in one or more places). Some of this information may stem directly from the fiction, but the information may also have to do with the roles that fictional characters play in the lives of those who think and talk about them, for instance the information that this afternoon was the last time anyone mentioned Lois Lane in your presence. The associated information may also have to do with the beliefs and other attitudes that some of the protagonists form about some of the others. If S has beliefs the content of which is information of this last kind, then she may want to convey such a belief to H, for instance by using one of the sentences in (3.28). She will be able to make us of the name *Lois* in any of these sentences because she is connected with the protagonist Lois through her *Lois*-labeled entity representation and her *Clark Kent*- and *Superman*-labeled entity representation will enable her to use the names *Clark Kent* and *Superman*. Furthermore, we may assume that her representation of the content of the beliefs she attributes involves attributing to Lois a $ClarkKent$-labeled and a *Superman*-labeled entity representation that, by S's own stipulation, are linked to her own entity representation that is labeled with both $ClarkKent$ and *Superman*. H, who may be assumed to have his labeled entity representations for Lois and Clark Kent/Superman can be expected to interpret S's use of the names in question the way he should and that she will expect him to.

That the belief which S expresses by uttering one of the sentences in (3.28) correctly captures the content of Lois' belief and that H will interpret her utterance in a manner that captures the belief she attributes to Lois correctly, cannot be argued on the basis of the little we have said here about the account of (Kamp 2021*b*). But going into the further details of the account that would be needed for such an argument would stand in no reasonable proportion to our aims here. Here we have only been concerned with an answer that could be given to the question how discourse participant can make coherent uses of fictional names in attitude attributions to fictional characters on the basis of their own name-labeled entity representations. And at the same time we also wanted to give an idea of how making such attributions to creatures of fiction is different from making attributions to real agents from our world and what is common to these two kinds of attributions.

Among the widely discussed cases of taking to be two what in reality is one are those discussed by Kripke in (Kripke 1979) – that of the hapless Pierre who doesn't realize that the dreadful place where he has washed up and knows by the name *London* is the same that was described to him in glowing terms by his friends in the France of his youth as *Londres*, and the case of a man called Peter, who thinks that Jan Paderewski the piano virtuoso and composer couldn't be the same person as the Jan Paderewski who was the first president of the Polish Republic. Kripke mentions this second case to show that the problems connected with the *London-Londres* example have nothing of importance to do with the morphological difference between *London* and *Londres*. That seems clearly right, at least as regards the way we can represent the mental states of those whom we assume are wrong in thinking that their representations of what is in fact one entity are the representations of two different entities. But there *is* a problem, both in the *London*/*Londres* case and the Paderewski case, about the use of names when reporting attitudes of the respective attributees Pierre and Peter. And for obvious reason the options in these two cases are not the same.

Let us begin with a closer look at the case of the unfortunate traveler Pierre. Suppose S and H both know about Pierre's misconception that what he refers to as *London* and *Londres* are different cities and that S wants to say something to H that has to do with that. Perhaps S thinks one should explain Pierre's misconception to him and has turned to H to ask him what might be the best way to do this. She might address H with something like this.

(3.30) Pierre thinks that Londres is different from London. He was told as a child that Londres is very beautiful and mentioned to me a few places that they showed him pictures of. Should we perhaps take him to a couple of those places?

Let us assume that S and H are inhabitants of London and native speakers of English. Then the use of *Londres* in the middle of this otherwise fully English piece of discourse is marked for that reason alone, and all the more so as *London*, the standard English denotation for Britain's capital, occurs in it as well. The markedness needs to be accounted for somehow, and given the content of (3.30) and the context that S and H share, it may be intuitively clear to H what this occurrence of *Londres* in (3.30) is meant to achieve. But even if it will be intuitively clear to H what S is trying to say, the way *Londres* is used in (3.30) does not seem felicitous. Better would be:

(3.31) When Pierre grew up as a child in France, they told him what a beautiful place 'Londres' was (as they referred to the place, being French), showing him pictures of some of its most attractive parts. Now he thinks that Londres must be a distinct place from London, which isn't surprising given the dismal area around the docks to which he has been confined since he has been here. Should we perhaps make it clear to him that he is wrong by taking him to a couple of those places they told him about?

In (3.31) the special role that *Londres* is playing in what S is saying is carefully prepared in the opening sentence. Once that has been done, it is much easier to accept the sentence with its second token – 'Now he thinks that Londres must be a distinct place from London' – as a legitimate expression of what it wants to say. And a similar effect can also be achieved by adding a suitable augmentative to it, as in *Londres, the name by which he first heard about London when he grew up as a child in France.* Adding augmentatives is one way in which we can make explicit that the augmented name is meant to play a role that it could not play (or not very well) without the augmentation

Precisely how to articulate what it is that makes (3.31) so much better than (3.30) isn't something we feel able to do. And we feel even less able to state general criteria that distinguish the good cases of using distinct names in attitude attributions to agents who use them as labels of entity representations that they mistakenly think represent distinct referents. But even in the absence of a more probing treatment of such cases, contrasting examples like (3.30) and (3.31) can serve to make us see that the use of such names for the purpose of pointing the interpreter to the different entity representations they label is subject to restrictions. In this regard the highly conventionalized use of *Hesperus* and *Phosphorus* among philosophers may well be an exception.

More or less inevitably, a similar predicament arises for a speaker S who wants to say something about attitudes of Peter, who has *Paderewski*-labeled entity representations for what he takes to two different Paderewski's, but which are anchored to a single person by that name. Saying things like 'Peter thinks that Paderewski is different from Paderewski', 'Peter thinks that Paderewski and Paderewski are different people' sound odd even in contexts where it may be clear what the speaker is getting at. The more natural way to say this would be some such thing as 'Peter thinks that Paderewski the piano virtuoso is different from Paderewski the politician'. The augmented name *Paderewski the piano virtuoso* has the property that the attributee Peter would recognize it as a possible label for one of his entity representations for Paderewski, but not as a label for his other entity representation; and likewise he would be able to recognize *Paderewski the politician* as a possible label for the second of those entity representations, but not for the first. In this way S can point H towards the first of those entity representations by using *Paderewski the piano virtuoso* as the one directly involved in the attribution she is making to him and point him to the second by using *Paderewski the politician.*

Name augmentations of the sort exemplified in our discussions of *London-Londres* and *Paderewski-Paderewski* appear to be a generally effective strategy in making attitude attributions to agents who fail to recognize that two of their entity representations stand for the same thing. It may be worth adding that such augmentations are used for other purposes as well. Another commonly employed kind of augmentation is where the name is modified by a phrase that denotes an era or period of time, as in *Paris anno 1910*, or *Paris in the early 20th century*, which can serve as the subject of predications that apply to those particular temporal parts of Paris' existence. (Thus 'Paris anno 1910/Paris in the early 20th century was the cultural and artistic center of the Western World' can be used to say the same thing as 'In 1910/the early 20th century Paris was the cultural and artistic center of the Western World'.) This is another example of where the augmented names serve to focus the interpreter on restricted sets of predications when compared with the non-augmented names.[21]

We sum up and conclude this discussion about misalignment of named entity representations and their referents with an emphatic repetition of what has been central to it: such misalignments pose two problems to a communication-theoretic treatment like MSDRT's:

(i) How do those who are aware of the misalignment represent it as part of their representation of the mental states of those who are misaligned?

(ii) What words is it appropriate to use in attitude attributions to those who are misaligned, and – this is a point that deserves further elaboration than we have given it here – how may that depend on what their audience does or does not know about the misalignment?

One important feature of the misalignments we have been dealing with is that – fortunately – they tend to be unstable: more likely than not the agent will find out eventually that there is just a single referent and accommodate his

[21]A further point that is relevant to the literature on the Paderewski example and that almost feels to be too trivial to bear mentioning is that many names, and especially names of people, are ambiguous in the sense that they are the names of more than one person. In fact, some names – family names in cultures like ours, are ambiguous by design: as a default new-borns get their last name from their father, or at least that of either one of their parents. Furthermore, there are in general not all that many first names between parents can choose, and in some countries this is a closed list.. As a consequence it is reasonably common, and anything but surprising, for even combinations of first and second name to be ambiguous in this sense. There no doubt are and have been any number of different people named 'John Williams' or 'Pierre Dubois' or 'Karl Schmitt' or 'Jan Paderewski'. Perhaps it is a little surprising to some of us when there two people in the public eye that are both called 'Jan Paderewski'. But that may still seem less unlikely than that a single person, whatever his name, accomplished all that the famous Paderewski did. So it isn't hard to see how Peter could have got into his misalignment, nor is it difficult to describe the relevant part of his mental state in MSDRT terms. There sometimes will be a problem about how to phrase the statements that attribute certain attitudes to persons with Peter's predicament. But that point is taken up in the main body of the text.

mental state accordingly. Given the assumptions about entity representation we have made, this accommodation must take the form of merging the two entity representations into one. But that is never possible without further adjustment. One belief that must be abandoned is that the two entity representations represented distinct entities (see (3.25)). But often further beliefs associated with the two entity representations one had will not be compossible when reconstrued as beliefs about one and the same entity. In such cases revision will be required of the total belief set as a whole. Such revisions form a special subclass of belief revision more generally. Seeing to what extent revisions of this special kind instantiate the general principles identified by the theory of belief revision[22] and what on the other hand is special about them, is a topic that to our knowledge has not yet been explored.

3.3.3 Taking Two to be One

The examples considered in Sections 6, 6.1 and 6.2 are all about the error of taking coreferential entity representations to represent different entities. But there is a converse error – turning information about distinct entities into a single entity representation, which one presumes to stand for just one thing, We do not know of any widely discussed example of this in the literature. But we suspect that some such cases have been experienced by most people. See if the following discussion will put you in mind of some such experiences you have had yourself.

Our case has to do with the name *John Williams*. During much of the second half of the last century the name *John Williams* was highly prominent within music. Those were days when information about people was far less often accompanied by pictures of them than today. That was so in particular for the musical scene: you could listen to people perform, or to the performance of their compositions, without ever having an idea of what they looked like. Under such conditions it was much easier to be misled into thinking that *John Williams* was both the man who conducted the Boston Pops and composed the music for Star Wars and the one who was a virtuoso classical guitar player. Implausible perhaps, and a little daft. But possible it was. We were the living proofs that it was.

It is clear how such misalignments should be described in terms of MSDRT: the agent has a single entity representation which is the condensation point of information that in actual fact stems from two distinct information sources.

[22]See for instance (Gärdenfors 1988), (Gärdenfors 1992), (Hansson 1999).

One effect of this is that the anchor set of such an entity representation is hopelessly corrupted, as it will contain some anchors linking it to the one John Williams and some other anchors linking it to the other John Williams. And having an anchor set with anchors puling it in opposite directions is hardly better than having no anchors at all.

Descriptions of the plight of someone like us by someone who knows better will typically make use of augmented names like *John Williams, the classical guitarist* and *John Williams, the composer of the music for Star Wars*. In this regard misalignments of the 'two are taken to be one' kind are no different from those of the 'one is taken to be two' kind. But when the agent comes to realize that there are two entities there instead of one, then what has to happen is quite different. The agent must now split his entity representation into two and that requires separating all the information in and associated with the old entity representation and the set of attitudes associated with it into the part that should go to the first of these new representations and the part that should go to the second (and perhaps also some third category of information that should be discarded altogether).

This is yet another type of belief revision. It too deserves to be looked into more closely, we think, both as compared with the revisions needed in the cleaning up of 'one is taken to be two' errors and in relation to general theories of belief revision.

With this we end our discussion of the use of names in attitude attributions.

Chapter 4

Introduction to MSDRT II: Model Theory

In our introduction to DRT we emphasized the importance of a model-theoretic semantics for DRT's DRS Languages. The models we developed were non-extensional along two dimensions: Indexed families of extensional models, where the indices are pairs $<w,t>$, consisting of a possible world w and a time t from the time structure of w. Models that are 'intensional' in this way will also be the models for DRS languages of MSDRT.

The general scheme for evaluating natural language sentences, discourses and texts for truth or falsity in these models will be as before too: The Logical Forms K of sentences, discourses and texts will be evaluated only at indices $<w,t>$ at which the sentence, discourse or text was uttered and where K was constructed as its Logical Form in the given context by a recipient with access to the relevant contextual information.

The models for MSDRT differ from those we presented in our introduction to DRT in that they must provide information about the mental states of agents. The models for any given language L must provide entities for each of the different sorts that L can refer to. For the DRS languages of MSDRT, which can refer to mental states through the MSDs that they allow as arguments of *Att*, this means that MSDRT models must provide, as part of their ontology, mental states of agents. And the mental states they provide must be such that the MSDs occurring in *Att*-Conditions can be assessed as correct or incorrect descriptions of them.

But what should the specifications of mental states in models be like? That is not an easy question. We will try to work our way towards an answer to it in steps.

First, we take it for granted at this point that the mental states of agents have propositional attitudes and entity representations among their constituents, just like the MSDs found in MSDRT's Logical Forms. After all, the idea that mental states are made up of propositional attitudes and entity representations was one of MSDRT's points of departure. Some recent developments of MSDRT assume that mental states have a more complex structure than that of mere sets of entity representations and propositional attitudes (e.g. (Kamp 2021*b*); also see the final part of this Introduction). But in the model theory we present in this section we stick with the simpler concept of mental states whose constituents are propositional attitudes, entity representations and nothing else.

The next question is how models should identify the propositional attitudes and entity representations that are the constituents of mental states. There are various options to be considered here, but it is helpful to set these aside for now and to turn to them later. We will do so in Section 6 of Ch. 4. For most of Chapter 4, however, we assume that the mental states of agents are MSDs, and thus of the same form as the descriptions of mental states found in MSDRT's Logical Forms for attitude attributions. For a model M for MSDRT this means that it must supply, for the relevant indices $<w,t>$, the mental states that the agents existing at those indices are in. That is, M must have a partial function MS_M (for 'mental states in M') which assigns to each appropriate pair $<<w,t>,a>$, consisting of an index and an agent a who is in a mental state at $<w,t>$, the mental state of a at $<w,t>$. As said, these mental states are assumed to take the form of (typically large) MSDs. But in addition they should also supply information about the entities represented by ERs in the MSD. The ERs of actual mental states will normally be causally connected to the entities they represent. There is no need for the model to say much about the nature of those causal relations. But we need it to tell us which entities are represented by which entity representations of each agent's mental states. To this end we assume that the function MS_M assigns pairs $<MSD,REF>$ to index-agent pairs $<<w,t>,a>$, where MSD is a's mental state at $<<w,t>$ and REF is a partial function from the set DIST(MSD) of all distinguished drefs of ERs in MSD to entities from M:

(4.1) (Definition of the *models for MSDRT*)

M is a *model for MSDRT* iff M is a pair $<M',MS>$, where

(i) M' is an IH model for DRT and

(ii) MS is a partial function from pairs $<<w,t>,a>$, where $<w,t>$ is an index from M' and a an agent from M', to pairs $<MSD,REF>$, where MSD is an MSD and REF is a partial function from the set DIST(MSD) of all distinguished drefs of ERs in MSD to entities from M'.

The distinction between models M and their first components M' corresponds to that between the DRS languages for MSDRT and the DRS languages for DRT whose extensions they are. For each language L for MSDRT let L' be the language for DRT whose DRSs are all those of L in which *Att* does not occur. Then L' is a language for DRT and M' is a model for L'. The verification definition for languages like L' can be found in the standard introductions to DRT; ((Kamp & Reyle 1993); see also the DRT introductions provided for the University of Texas seminar that prompted writing the present MSDRT introduction.)

The main task of Chapter 4 is to formulate a correctness definition for MSDRT's Logical Forms for attitude attributions in models. Given what these Logical Forms are like, this boils down to the verification conditions for DRS Conditions of the form 's: $Att(a,MSD,\text{Links})$'. And the crucial question for the verification conditions of an *Att*-Condition is the relation between its argument MSD and the mental state of the agent (represented by the dref 'a' in the second argument position of *Att*) at the relevant evaluation index $<w,t>$. But before we get to the definition of the verification conditions and the relation between MSD and the agent's actual MSD at $<w,t>$, we will have to do a fair bit of preparatory work, which has to do with the model-theoretic semantics of MSDs: What are the denotations in a given model M of MSDs and their constituents? So long as an MSD does not contain any content-specifying DRSs that contain *Att*-Conditions in their turn, this last question is one that involves M only in its capacity of a model for the underlying DRS language of DRT (that part of the given DRS language of MSDRT in which there are no occurrences of *Att*-Conditions). In other words, the denotations in question are determined by the model M' of Definition (4.1). (This also means that Logical Forms for iterated attributions – attributions of thoughts whose contents are attitude attributions – can be dealt with only once the semantics of non-iterative attributions is in place; see Ch. 4, Section.3.2). Until then the DRS denotations we will be dealing with are denotations of DRSs belonging to the underlying DRS language of DRT.

4.1 Denotations for MSDs and their Constituents

Since we are assuming that the mental states of agents in models are MSDs, the questions about MSD semantics that we need to clear before we are able to state our correctness definition for attitude attributions apply to the actual mental states of agents just as they apply to the MSDs occurring as arguments of the *Att*-Conditions that are part of Logical Forms for mental state attributions. It will be only in Ch. 4, Section 5 that we will look at alternatives to the assumption that mental states are MSDs.

4.1.1 Live and dormant anchors of entity representations. Doxastic Import

We start with a couple of issues related to entity representations. The ERs found in MSDs have both a descriptive and a causal side to them. As we have seen, descriptive information is found in both the second and the third component of an ER. The second component is a DRS from the underlying DRS language of DRT. For the entity representations belonging to the mental states of agents in a model M the contents of these DRSs will be identified as M-propositions, as defined by the model theory for DRT. What to do with the descriptive contents of the anchors of entity representations may be less obvious. This has to do with an aspect of anchors that we have barely touched upon so far: to what extent do individual anchors retain their presence in the consciousness of the agent after they have been introduced? On the one hand we might want to say that current awareness of an anchor on the part of the agent isn't a necessary part of its belonging to the anchor set of the given ER. Once the anchor has been introduced as witness to some given causal relation between the ER and the entity represented by it, the anchor is and remains part of the ER, irrespective of whether it has dropped below the agent's awareness threshold. On the other hand, however, anchors of which the agent is aware – in the form in which she was when they were first introduced or in some modified form imposed by the passage of time – may play a decisive part in her reasoning about the represented entity, for instance because of the information it gives the agent about how she first became acquainted with the entity. Since the distinction between anchors above and below the agent's awareness threshold is clearly relevant to the roles that the entity representations in question can play in the agent's cognition, it should be part of our theory. To this end we refine the formal notion of an ER by assuming that the anchor sets of ERs are divided into two parts, $\mathcal{CA}$, the *cognitively accessible* anchors, and $\mathcal{DD}$, the *dead or dormant* anchors.[1]

The distinction between active and dormant anchors is evidently important or what cognitively relevant content is associated with an ER at any index $<w,t>$: only the $\mathcal{CA}$ anchors are relevant for this. That is, the *total active descriptive content* of an ER at index $<w,t>$ is given by the DRS that is the merge of the descriptive DRS in second position and the $\mathcal{CA}$ anchors in the anchor set. We assume that agents are always in a position to form a belief with this merge DRS as content specification. We refer to this belief as *the doxastic import of* the ER *at* $<w,t>$). We will assume occasionally that the doxastic import of an ER in an agent's mental state is a constituent of that state as well.

The doxastic import of an ER is available irrespective of whether there is an entity that the ER represents. (it is just that when there is no represented entity, the doxastic import will normally be false.) Furthermore, it is always the doxastic import of an ER that captures its cognitive significance – the role that the ER plays in the agent's practical reasoning, in her planning and decision making and so on. This cognitive significance is detached from any actual properties of the entity that is represented by the ER but to which the agent has no access. And that applies equally to all ERs from an MSD; each of them has its own doxastic import. If we replace in a given MSD every ER by its doxastic import, we obtain the *doxastic import of* the MSD. The doxastic import of an MSD MSD is an MSD MSD' that consists entirely of Propositional Attitudes. In MSD' all content specifications K of Propositional Attitude constituents express non-singular contents. This is true also for those K whose contents expressed singular propositions in MSD, because the ERs of the distinguished drefs they contain as arguments properly represent. In MSD' these drefs now occur in the Universes of the doxastic imports that replace their ERs. In MSD' referential dependence takes the place of singularity. (Compare our very first two examples of MSDs, (2.3) and (2.4).) It is because the doxastic import of her mental state is crucial to the agent's cognition rather than her mental state itself (with its partly causally determined identity and structure) that referential dependence is such a crucial feature of mental representations and their mental manipulations.

[1]It is possible to distinguish between 'dead' anchors, which have been irretrievably wiped from the agent's mind, and 'dormant' anchors, which are still represented somewhere in the agent's cognitive system and which she might still retrieve under the right conditions. We see not good grounds, however, for building this distinction also into our formal definition of ERs.

4.1.2 Entity representations as arguments in the Content Specifications of Propositional Attitudes. Questions of Logical Ontology

A central assumption of MSDRT is that entity representations can provide argument terms (viz. their distinguished drefs) for the predications occurring in the content specifications of propositional attitudes. When an entity representation provides such an argument term, and when it is non-defective (it does represent an entity **d**), then the content of an attitude to which it provides the argument term will be a singular proposition about **d** that the entity representation represents. When all the entity representations that provide argument terms in the predications of a content specification properly represent, then the content specification will express a well-defined content. And when, moreover, the content specification K is referentially independent, then this content will be a proposition that is singular with respect to each of the entities involved – that is: with respect to each of the entities that are represented by entity representations whose distinguished drefs occur as arguments somewhere in the Conditions in the Condition Set of K.

Content specifications occur in our MSDs in three different kinds of positions: (a) as the content specifications of Propositional Attitude constituents; (b) as second components of entity representations; (c) as anchors of entity representations. (b) and (c) are taken care of in the doxastic imports of ERs and do not need separate consideration here. What matters are the DRSs used as content specifications of Propositional Attitude constituents.

Let K be the content specification of some Propositional Attitude constituent. The drefs occurring in the DRS Conditions of K can be classified as follows. First, such a dref can be bound within the Condition itself in which it occurs. This is possible when the Condition in which the dref occurs is a complex DRS Condition – a negation, disjunction. conditional, Duplex Condition etc. These drefs do not require special attention; their role is accounted for by the model-theoretic semantics for DRT that is part of the model theory for MSDRT. (See the introductory part of Chapter 4.) But the free occurrences of drefs in Conditions of K – those that are not bound within the Condition itself – do require attention. (4.2) gives the four-fold classification.

(4.2) Suppose that K is the content specification of a Propositional Attitude constituent of an MSD MSD that occurs as third argument of an *Att*-Condition in some larger DRS K' and that x is a dref with an occurrence in the Condition Set of K. Then there are the following four possibilities for how x is 'bound' in K':

(i) x is bound within K;

(ii) x is bound higher up in K';

(iii) x is the distinguished dref of an ER in MSD;

(iv) x belongs to the Universe of some other Propositional Attitude constituent of MSD.

The distinctions in (4.2) can be used to distinguish between different content specifications. The simplest case is that of a DRS K in which all drefs with free occurrences in the Condition Set of K are of type (i). This is the case of a 'self-contained' DRS.[2] In the intensional model theory for DRT such DRSs determine propositions in intensional models, where the propositions of an intensional model M are the subsets of its index set. So in the IH models adopted here, propositions are sets of world-time pairs. An example of such a content specification is that of the belief constituent of (2.3), repeated here:

$$\text{(2.3)}\quad \left\{\begin{array}{l} \left\langle BEL, \begin{array}{|c|}\hline x\quad s_1\quad s_2 \\ \hline n\subseteq s_1\quad n\subseteq s_2 \\ s_1\text{: gold-coin'(x)} \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline\end{array}\right\rangle \\[2em] \left\langle DES, \begin{array}{|c|}\hline s_3 \\ \hline n\subseteq s_3 \\ s_3\text{: have'}(i,x) \\ \hline\end{array}\right\rangle \\[2em] \left\langle INT, \begin{array}{|c|}\hline t\quad e \\ \hline n\prec t\quad e\subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline\end{array}\right\rangle \end{array}\right\}$$

In an IH model M the belief DRS K_{BEL} of (2.3) denotes the proposition consisting of those indices $<w,t>$ such that there is an assignment for the Condition Set of K_{BEL} in M at $<w,t>$ (where an assignment here is a function from the Universe of K_{BEL} to entities from M). As we will see in the next section, propositions will not be among the central semantic notions of

[2]In (Kamp & Reyle 1993) such DRSs are called *proper DRSs*.

our semantics for MSDs. But they will still play a derivative role.

An example of a content specification that involves drefs of both type (i) and type (iii) is that of the belief constituent of the following example (inspired by the work of Sainsbury, where examples involving Ponce de Léon occur in several places (Sainsbury 2005), (Sainsbury & Tye 2012), (Sainsbury 2018).

(4.3) Ponce de Léon believed that there was a rejuvenating spring in Florida; and he wanted to find it.

(4.4) is an MSD for the relevant part of Ponce de Léon's mental state at the time that the attitude report (4.3) talks about.

$$(4.4)\quad \left\{ \begin{array}{l} \left\langle [ENT, f], \begin{array}{|c|}\hline \\ \hline \text{Named}(f,\ \textit{Florida}) \\ \hline\end{array}, \mathcal{K} \right\rangle \\ \\ \left\langle BEL, \begin{array}{|c|}\hline b \\ \hline \text{rejuvenating-spring'}(b) \\ \text{in'}(b,f) \\ \hline\end{array} \right\rangle \\ \\ \left\langle INT, \begin{array}{|c|}\hline t\ \ e \\ \hline n < t\ e \subseteq t \\ e\text{: find'}(i,b) \\ \hline\end{array}\right\rangle \end{array} \right\}$$ [3]

The Condition Set of the content specification K_{BEL} of the belief constituent of (4.4) has two free occurrences of drefs, a type (i) occurrence of b and a type (iii) occurrence of type f. The proposition denoted by K_{BEL} in M is

[3]Ponce de Léon (1474-1521) was one of the Spanish Conquistadores and the first Governor of Puerto Rico. The story that he believed there was a 'Found of Youth' in Florida is apocryphal, though he did organize an expedition there. According to the MSD in (4.4) Ponce de Léon had an entity representation for Florida, and his belief that there was a rejuvenating spring there is represented as a belief about the entity represented by this ER. It apopears to be an established fact that Florida got its name only when Ponce de Léon called it that upon arrival there. If so, then the ER of (4.4), which includes the information that Florida went be the name *Florida* is inaccurate when (4.4) is a description of Ponce de Léon's state of mind before the expedition to Florida. We have taken this liberty for illustrative purposes.

the set of those indices $<w,t>$ of M at which K_{BEL} is verified by a suitable assignment g. But what is it for g to be a suitable assignment in this case? First, the Domain of g must consist of the two drefs f and b. Second, $g(b)$ should be some suitably chosen entity **b** from M, and thirdly $g(f)$ should be the entity represented by the entity representation of (4.4). But what is that entity?

Answer: This now depends on whether (4.4) is the actual mental state of Ponce de Léon at the relevant time (or some part thereof) or rather an MSD offered as description of this or some other agent's mental state. In the first case, the entity is given by the reference function REF_M; that is, $g(f)$ must be $REF_M(f)$. But when (4.4) is part of the Logical Form of an attitude attribution, then the only way to make sense of the denotation of K_{BEL} is by reference to the mental state of the attributee at the relevant index. What this means for the correctness of MSDs and the *Att*-predications containing them as parts of Logical Forms for attitude attributions can be made explicit only in Section 3.2. At this point no more can be said about this aspect of the denotations of content specifications than that they depend on whatever entities are assigned to the distinguished drefs that occur in them.

The main challenge we are facing is the semantics for content specifications that contain drefs of type (iv). MSDs with content specifications with such dref occurrences were the original motivation for MSDRT and the central problem that was perceived from the start for its model-theoretic semantics. The way we will meet that challenge is in essence still the same one. It makes use of the formal tools developed as part of Dynamic Semantics in the narrower sense of that term. (That sense in which it refers to the 'non-representational' approaches to problems of discourse anaphora first developed by Groenendijk and Stokhof.[4] These tools and their application to MSD semantics are the topic of the next section.)

[4]See (Groenendijk & Stokhof 1990), (Groenendijk & Stokhof 1991). The early work by Groenendijk and Stokhof on Dynamic Semantics was in part motivated by DRT: to develop an account of pronominal anaphora that would make the same predictions as the first versions of DRT, and also by Heim's File Change Semantics (Heim 1982,1988), (Heim 1983), which makes the same predictions as DRT, but without DRT's representational commitments – that is, without making use of representational structures like DRSs. Both in its conceptualization and in its technical implementation Dynamic Semantics is closer to File Change Semantics than to DRT. The version of Dynamic Semantics that is the most direct inspiration for the use we are making of its ideas and formal tools can be found in joint work of Groenendijk, Stokhof and Veltman, (Groenendijk, Stokhof & Veltman 1996*a*), (Groenendijk et al. 1996*b*).

4.2 Using Dynamic Semantics in the Model Theory for MSDRT

Examples of a content specification with dref occurrences of type (iv) are the DRSs in the desire and intention constituents of (2.3), our first illustration of an MSD, which we repeat once more.

$$(2.3)\quad \left\{ \begin{array}{l} \left\langle BEL, \begin{array}{|c|}\hline x\quad s_1\quad s_2 \\ \hline n \subseteq s_1\quad n \subseteq s_2 \\ s_1\text{: gold-coin'(x)} \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array} \right\rangle \\ \\ \left\langle DES, \begin{array}{|c|}\hline s_3 \\ \hline n \subseteq s_3 \\ s_3\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\ \\ \left\langle INT, \begin{array}{|c|}\hline t\quad e \\ \hline n \prec t\quad e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

The content specifications of the desire and the intention constituent of (2.3) referentially depend on the content specification of the belief constituent by virtue of their sharing the dref x, which occurs in the Condition Sets of the desire and the intention DRS, but is bound through its occurrence in the Universe of the belief DRS. Such instances of referential dependency cannot be dealt with in the manner of DRT, by lumping the contents of the desire and intention constituents together with that of the belief constituent. The mental life of an agent A whose mental state is correctly described by (2.3) depends crucially on her ability to keep the three contents separate, since they play distinct roles, indicted by their Mode Indicators, in A's mental processes.

A natural way to keep the contents of belief, desire and intention in (2.3) separate is provided by 'Dynamic Semantics' in the narrower use of this term introduced towards the end of Section 1.2 of this chapter. The part of Dynamic Semantics that we are going to make a central use of is its replacement of the notion of a proposition by a pair of notions, that of an *information* *state* and that of a *context change potential* (*ccp*). In the model theory for MSDRT we will present in the remainder of Chaprer 4, information states and context change potentials replace propositions as the central players.[5]

To see how this can be done, we first consider the semantics of a DRS like the belief content DRS K_{BEL} of (2.3). This belief DRS is self-contained in the sense explained in Section 1.2: every dref that has one or more free occurrences in its Condition Set also occurs in its Universe (and is bound thereby). Here is the definition of the *propositional function* defined by the Condition Set of K_{BEL} *in* a model:

For each model M the Condition Set of K_{BEL} defines a propositional function in M. This is the function whose arguments are assignments of entities in M to the drefs in the Universe of K_{BEL} and whose values are propositions of M, where the propositions of M are identified with sets of indices $<w,t>$ from M. For instance, let g be the assignment to the drefs in the Universe of K_{BEL} that maps s_1 to the state $\mathbf{s}_1$ of M, s_2 to the state $\mathbf{s}_2$ of M and x to the individual $\mathbf{d}$ of M. The propositional function determined in M by the Condition Set of K_{BEL} maps g to the proposition which says of each index $<w,t>$ at which K_{BEL} can be evaluated[6] that $\mathbf{s}_1$ and $\mathbf{s}_2$ hold in w at t, that $\mathbf{s}_1$ is a state to the effect that $\mathbf{d}$ is a gold coin in w at t and that $\mathbf{s}_2$ is a state to the effect that $\mathbf{d}$ is lying in front of the agent, represented as i, in w at t. In other words, this is the proposition in M that consists of all the pairs $<w,t>$ for which these conditions are satisfied.

The self-containedness of the belief DRS of (2.3) makes it possible to assign this DRS its existential DRT semantics, according to which it expresses the

[5]The presentation that follows is in essence that first described in (Kamp 2003), the German translation of n English manuscript that is still available only in this translation. This publication does not seem to have received much attention (notwithstanding the exemplary quality of the translation, by Ulrike Haas-Spohn). A reason may be that German isn't easily accessible to many members of the international philosophy and linguistics communities. The presentation here differs from that in (Kamp 2003) in some respects, among them in its more explicit treatment of the contributions made by entity representations.

[6]Quantified statements about evaluation of DRSs at indices of models are generally subject to such restrictions because evaluation isn't always possible. In the present instance this might be because at a given index $<w,t>$ the agent whose mental state is described in (2.3) does not exist. So there is no mental state for the agent at that index in terms of which the belief should have to be evaluated. (This of course then holds equally for the other two components of (2.3).) There can be all sorts of reasons that prevent a DRSs from evaluation at certain indices of certain models.

proposition that there is some way of assigning to the drefs in its Universe entities from M that satisfy the propositional function. But note that this existential proposition can be derived from a more informative semantic object defined by K_{BEL}: the function SAT_{BEL}, which maps each relevant index $<w,t>$ onto the set consisting of all assignments of entities to the drefs of the DRS Universe that satisfy all the Conditions in the Condition Set at $<w,t>$. The proposition expressed by the belief DRS then consists of all pairs $<w,t>$ such that $SAT_{BEL}(<w,t>)$ is non-empty. It is in this sense that the notion of a proposition has now become a derivative notion. Functions like SAT_{BEL}, from which propositions can be derived in the way shown, are called *information states*.

The information state SAT_{BEL} determined in M by the belief DRS K_{BEL} of (2.3) has the property that all assignments occurring in it (as members of the sets $SAT(<w,t>)$ for various indices $<w,t>$) have the same domain, consisting of the drefs s_1, s_2 and x). This set is nothing other than the Universe of K_{BEL}, and the same will be true also for information states determined by other self-contained DRSs in the manner shown above. All such information states have 'domain uniformity' in this sense. Domain uniformity is a property that we will assume for information states in general, irrespective of whether they are determined by DRSs. This feature is built in to the definition of 'information state in M given below in (4.5.b).

(4.5) a. Let M be an IH model, D a set of discourse referents. An *information state of M with argument set D* is a function SAT defined on the set of indices of M such that for each index $<w,t>$, $SAT(<w,t>)$ is a set of assignments from D to entities from M.

b. SAT is an *information state of* M iff there is some set D of drefs such that SAT is an information state of M with argument set D.

c. For any information state SAT of M the unique set D such that SAT is an information state of M with argument set D we refer to D as '$ArgD(SAT)$'.

The set of information states in a model M that are determined by self-contained DRSs is in general a proper subset of the set of all information state in M. But in the model theory for MSDRT this subset is of primary importance, and deserves some special notational convention. Suppose that K is a self-contained DRS, M a model and that SAT is the information state determined by K in M. Then we denote SAT also as '$[[K]]_M$'. (4.6) gives the formal definition of $[[K]]_M$ and its extension to a larger class of content specifications.

We have seen how self-contained DRSs determine information states in models and that this is so in particular for self-contained content specifications in MSDs. The official definition is given below in (4.6.a). This definition can be extended to content specifications K with the property that the free drefs of their Condition Sets are of types (i), (ii) or (iii) in the sense of Definition (4.2). (That is, the drefs either belong to the Universe of K or bound higher up in the DRS of which the MSD to which K belongs is a part or are the distinguished drefs of some ER in the MSD.) The extension is given in (4.6.b).

(4.6) a. Let K be a self-contained DRS, M be an IH model for a DRS language to which K belongs. The *information state denoted by* K in M, $[[K]]_M$, is the function SAT defined on the set of indices of M such that for each index $<w,t>$ of M, $SAT(<w,t>)$ is a set of assignments f from the Universe of K to entities from M such that f verifies in M at $<w,t>$ all Conditions in the Condition Set of K.

b. Let K be the content specification of a constituent $<MOD,K>$ of an MSD MSD that is part of a larger DRS K' and assume that the free drefs of the Condition set of K are of types (i), (ii) and/or (iii) in the sense of Def. (4.2). Let f be an assignment of entities from M to the drefs in K of types (ii) and (iii). Then the *information state determined in* M *by* K *as part of* MSD *relative to* f, $[[K]]_{M,MSD,f}$, is defined as follows:

For each index $<w,t>$ of M, $([[K]]_{M,MSD,f})(<w,t>)$ consists of assignments g such that (i) $\text{Dom}(g) = \text{Dom}(f) \cup U_K$ and (ii) g verifies all Conditions in the Condition Set of K at $<w,t>$ in M.

The instances of (4.6.b) that will be of most direct relevance later on are those where f assigns to each distinguished dref in its domain the entity represented by its ER. We will be able to make this precise only in Section 3.2.

In our discussion above of the information state determined by K_{BEL} in M we spoke of the 'propositional function' determined by the Condition Set of K_{BEL}. In our formal definition (4.5) of the notion of an information state, however, propositional functions are no longer mentioned. There is no conflict here, since the propositional functions – functions from tuples of entities to propositions; that is, functions from tuples of entities to sets of indices – can be straightforwardly recovered from the information states SAT defined in (4.5) (and so, therefore, can the propositions that are derivable from the propositional functions by existential quantification over the satisfying assignments). Throughout the remainder of Section 2 and in Sections 3 and 4 the propositional functions determined by Condition Sets won't play a role. But we will return to them in Section 5.

4.2.1 Context Change Potentials

So far we have just considered the semantics of K_{BEL} and other self-contained content specifications of Propositional Attitudes in MSDs. For these the denotations in a model M are, we have just seen, information states in M. When the content specification of a Propositional Attitude is not self-contained, however, its denotation cannot be an information state. The denotations of such DRSs must be *context change potentials*, or *ccp*s for short. *ccp*s are functions from information states to information states; but to specify more precisely what they are we need answers to three questions: (i) Which functions are determined by which content specifications? (ii) For which information states are those functions defined, and (ii) Why are these functions the denotations we want?

As a first step towards answering these questions, consider the desire content K_{DES} of (2.3). This DRS is not self-contained; its Condition Set has a free occurrence of the dref x that does not occur in its Universe. This means that the propositional function defined by this Condition Set cannot be applied to the entities specified by an assignment to the drefs in the Universe of the DRS; we also need a value for x and an assignment that is only defined on the drefs in the Universe will not provide a value for x. The value for x has to come from somewhere else. But where should we get it?

To answer this question we return to our classification in (4.2) of the different ways in which the free drefs in the Condition Sets of non-self-contained content specifications for Propositional Attitudes in MSDs can be 'bound' elsewhere. For easier reading we repeat this classification here.

(4.2) Suppose that K is the content specification of a Propositional Attitude constituent of an MSD MSD that occurs as third argument of an *Att*-Condition in some larger DRS K' and that x is a dref with an occurrence in the Condition Set of K. Then there are the following four possibilities for how x is 'bound' in K':

(i) x is bound within K;

(ii) x is bound higher up in K';

(iii) x is the distinguished dref of an ER in MSD;

(iv) x belongs to the Universe of some other Propositional Attitude in MSD. constituent of MSD.

Of the three possibilities in (4.2) it is the fourth that we want to concentrate on. (The possibilities (ii) and (iii) have to do with aspects of the over-all structure of K' and will be dealt with later, in Section 3.2. And Drefs of type (i) are of no direct concern either; if all the free drefs in the Condition Set were of this type, then K would be self-contained and its denotation an information state. We will see later on that this is also the case when the free drefs are all of either one of types (i)-(iii).) It is free drefs of type (iv) that make it necessary to adopt denotations with the more complex structure of *ccp*s.

The problem that x presents for the semantics of K_{DES} in (2.3) is that intuitively it gets its value from the specification K_{BEL} of the belief constituent. If the belief is true at some index $<w,t>$ in M, then there will be some assignment f of entities of M to the drefs in the Universe of K_{BEL}. f will assign some value to x and this value can then also serve as value for x in the semantic evaluation of K_{DES} in M at $<w,t>$. To this end it must be combined with values for the other drefs with free occurrences in the Condition Set of K_{DES}.

To see more easily what the different roles are that the two types of free dref occurrences play in this example – the role of s_3, which is of type (i), and the role of x, which is of type (iv) – let us have a look at a discourse that we first considered in our discussion of the Attitudinal Hierarchy in Section 2.2.1 of Chapter 2. This discourse, given in Ch. 2 as (2.34.a), is given below.

(2.34.a) John knows that a friend of his owns a house in Cedar Park. He believes that she rents it out to some lawyer. He hopes the lawyer takes good care of it.

A simplified Logical Form for (2.34.a), in which all temporal information has been suppressed (like we did in the Logical Form we gave for (2.34.c) in Ch. 2, Section 2.2.1, is shown in (4.8).

(4.7) John knows that a friend of his owns a house in Cedar Park. He believes that she rents it out to some lawyer. He hopes the lawyer takes good care of it.

$$
(4.8)\quad \left\{
\begin{array}{l}
\left\langle KNOW, \begin{array}{|c|}\hline f \quad h \\ \hline \text{friend-of'}(f,i) \quad \text{house'}(h) \\ \text{own'}(f,h) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline l \\ \hline \text{lawyer'}(l) \quad \text{rent'}(f,h,l) \\ \hline \end{array} \right\rangle \\[2ex]
\left\langle HOPE, \begin{array}{|c|}\hline \\ \hline \text{take-good-care'}(l,h) \\ \hline \end{array} \right\rangle
\end{array}
\right\}^{7}
$$

Let M be a model for this language. Then the know-DRS K_{KNOW} of (4.8) will determine an information state SAT_{KNOW} in M. The assignments in the sets $SAT(<w,t>)$ for varying indices $<w,t>$ will be assignments to the set $\{f,h\}$ of the two drefs f and h in the Universe of K_{KNOW}. The belief DRS K_{BEL} of (4.8) cannot determine an information state because the drefs f and h have free occurrences in its Condition Set but don't occur in its Universe. The content specification K_{BEL} of the belief in (4.8) determines a *ccp*, like the specification of the desire constituent of (2.3). (Much the same is true also for the content specification of the hope constituent of (4.8), with the only difference that this specification referentially depends both on the knowledge constituent and the belief constituent. We will turn to cases of referential dependence on more than one MSD constituent later.)

How the roles of f and h differ from that of l, let us consider what ought to be the result when the *ccp* $[[K_{BEL}]]_M$ determined in M by the K_{BEL} of (4.8) is applied to the information state $[[K_{KNOW}]]_M$ determined in M by (4.8)'s knowledge specification. This will also enable us to make a first step towards a general definition of how content specifications determine *ccp*s,

Our proposal: The application modifies $[[K_{KNOW}]]_M$ in two ways, stated in (4.9.a) and (4.9.b).

(4.9) Let $[[K_{KNOW}]]_M$, $[[K_{BEL}]]_M$ be as described above and let SAT' be the information state that results from applying the former to the latter. Then:

[7] As noted in footnote 14, Ch.2, Section 2.2.1, the Mode Indicator 'KNOW' is a purely doxastic one with a high degree of certainty.

a. For arbitrary indices $<w,t>$ of M the assignments in the sets $SAT'(<w,t>)$ are functions whose domain is $\{f,h,l\}$.

b. For any index $<w,t>$ and assignment g to $\{f,h,l\}$, $g \in SAT'(<w,t>)$ iff there is an f such that (i) $f \in SAT(<w,t>$, (ii) $f \subseteq g$, and (iii) g verifies K_{BEL} in M at $<w,t>$.

Exercise: Show that the proposition derived from SAT' entails the proposition derived from SAT.

(Note in connection with this exercise that $SAT'(<w,t>)$ may be empty even when $SAT(<w,t>$ is not: Although $f \in SAT(<w,t>)$, there may be no g defined on $\{f,h,l\}$ such that $f \subseteq g$ and g verifies K_{BEL} in M at $<w,t>$. So for no extension g of f, $g \in SAT'(<w,t>)$. And this may so for every $f \in SAT(<w,t>)$. On the other hand, when $SAT(<w,t>)$ is empty, then of necessity $SAT'(<w,t>)$ will be empty too.)

From this description of how the *ccp* determined by K_{BEL} in (4.8) is applied to the information state determined by its knowledge DRS K_{KNOW} it should be clear what the result ought to be when this *ccp* is applied to other information states, and what the information states are to which application is possible:

(4.10)Let M be as above. The *ccp* determined in M by the belief-DRS of (4.8) is a partial function from information states in M to information states in M. This function is defined for those information states SAT in M such that $ArgD(SAT)$ contains the type (iv) drefs f and h and does not contain the type (i) dref l. And for any such information state SAT the result of applying the *ccp* determined by K_{BEL} to it results in an information state SAT' that is obtained from it in the way described in (4.9).

Just as (4.5) is a general definition of the information states in a model M, independently of whether these can be described by our DRS languages, so we also want a general, representation-independent definition of the *ccp*s of M. This general notion should conform, however, to the principle that holds for all *ccp*s that are the denotations of referentially dependent content specifications: there is a fixed set D of drefs such that when SAT is an information state in M for which a given *ccp* CCP is defined, then $ArgD(SAT) \cap D = \emptyset$ and if SAT' is the result of applying CCP to SAT, then $ArgD(SAT') = ArgD(SAT) \cup D$. (N.B. when CCP is the denotation of a referentially dependent content specification K, then D is the Universe of K.) The definition of the set of *context change potentials in* M includes this constraint.

(4.11)a. Let M be an IH model, $PRES$, D a pair of two disjoint sets of drefs. Then a *ccp of M for $PRES$* and D is a partial function from information states in M to information states in M which is defined on those and only those information states SAT of M such that $PRES \subseteq ArgD(SAT)$ and $ArgD(SAT) \cap D = \emptyset$. Furthermore, if $SAT' = ccp(ArgD(SAT'))$, then $ArgD(SAT') = ArgD(SAT) \cup D$.

b. CCP is a *ccp of M* iff there are sets $PRES$ and D such that CCP is a *ccp of M* for $PRES$ and D.

Note that it follows from (4.11.a) that for any *ccp* CCP of M the sets $PRES$ and D such that CCP is a *ccp* of M for $PRES$ and D are uniquely determined. Moreover (and repeating more or less the motivation for treating content specifications K of Propositional Attitude constituents of MSDs as denoting *ccp*s), when CCP is the denotation in M of such a content specification K, then D is the Universe of K and $PRES$ is the set of drefs with free occurrences in the Condition set of K.

We conclude this section with what may seem a trivial observation about formal cosmetics. But trivial as it may be, it points to a connection between information states and context change potentials that is not immediately visible from the presentation we have given and illuminates the use we are making of these notions; and that is why we include it here. The observation is that information states can be thought of as *ccp*s of a special kind. From this perspective the denotations of all Propositional Attitude constituents of MSDs are *ccp*s. It is just that some of them are of a special kind, which enables them to shore up the hierarchical semantic structure of the MSDs to which they belong.

We first define the *Merge* of two information states. Let M be a model and let SAT and SAT' be information states in M.

(4.12)The *Merge* of SAT and SAT', $SAT \oplus SAT'$, is the information state SAT'' such that for each index $<w,t>$ of M $SAT''(<w,t>) =$ $\{h\text{: } (\exists f)(\exists g)(f \in SAT(<w,t>)\ \&\ g \in SAT'(<w,t>)\ \&\ h = f \cup g)\}$.

Note that if SAT and SAT' are the denotations $[[K]]_M$ and $[[K']]_M$ of DRSs K and K', then $SAT \oplus SAT' = [[K \oplus K']]_M$, where $K \oplus K'$ is the DRS merge of K and K'.

We now associate with each information state in M a *ccp* as follows:

(4.13) Let SAT_0 be an information state in M. We refer to $ArgD(SAT_0)$ as D_0 and to the *ccp* that is to be associated with SAT_0 as ccp_0. The definition of ccp_0 is as follows:

(i) ccp_0 is defined on all and only the information states SAT in M such that $ArgD(SAT)$ is disjoint from D_0.

(ii) Let SAT be such an information state. We refer to $ccp_0(SAT)$ as SAT'. SAT' is defined by: $SAT' = SAT \oplus SAT_0$.

As the definition shows, the *ccp* ccp_0 associated with an information state SAT_0 is maximal in the sense that ccp_0 is defined for *all* information states SAT such that $ArgD(SAT) \cap ArgD(SAT_0) = \emptyset$. This includes in particular the *trivial information state* SAT_T defined by:

$$\text{for all } <w,t>,\ SAT_T(<w,t>) = \{\emptyset\}.$$ [8]

It is easily verified that $CCP_0(SAT_T) = SAT_0$: the information state with which CCP_0 is associated is one of its function values. In view of this it is possible to think of the *ccp* associated with an information state as just another representation of that very information state.

This also makes it possible to define the denotations of MSD constituents whose content specifications are self-contained DRSs as *ccp*s. By this decision, all content specifications of Propositional Attitude constituents of MSDs now have *ccp*s as denotations, whether they are self-contained or not. But when we choose to do so, we must not forget that the unification we achieve in this way is a superficial one. The difference between the *ccp*s that return corresponding information states as values and those that do not is still there; and that difference is crucial.

4.2.2 Context Change Potentials and Primary Information States

This gives a complete characterization of the *ccp* denoted by referentially dependent content specifications of Propositional Attitude constituents of MSDs with the property that the drefs with free occurrences in their Condition Sets are either of type (i) or of type (iv). There is a complication, though, that we hinted at above, but haven't as yet addressed: A content specification may referentially depend on more than one other specification; and it is also possible that one or more of those specifications referentially depend in their

[8] Here $\emptyset$ is to be understood as the 'empty function', the one and only function whose domain is empty.

turn on yet others. One example of this is the hope constituent of (4.8). Its content specification K_{HOPE} depends on K_{KNOW} in virtue of h and on K_{BEL} in virtue of l. And K_{BEL} depends in its turn on K_{KNOW} by virtue of h and f.

Before we address the technical problems connected with such multiple dependencies, first a word about the psychological significance of referentially dependent constituents of mental states. For now just one example. Consider again the dependence of the desire constituent in (2.3) on its belief constituent. According to our proposal the denotation of its content specification in model M is the *ccp* described in (4.10). But what could be the psychological significance of a *ccp*? The traditional view of the sense in which attitude attributions tell us something about the mind of the attributee is that they attribute to her *propositions*, to which she has various propositional attitudes and that play their parts in her mental life in ways that are determined by those attitudes. For instance, propositions believed will serve as premises in her practical reasoning; and so can propositions desired, often in conjunction with propositions believed. We too regard this view as basically correct, though the question must be raised at some point how finely individuated the propositions should be from and to which we human agents reason. It seems to us a plausible conjecture that human reasoning often involves mental representations that are not too different from linguistic representations and that in some cases they are the very sentences that we also use when we do our reasoning in public. In the intensional modeling of mental processes that is the point of departure for the model-theoretic semantics for MSDRT presented in Chapter 4, propositions in our current sense, viz. sets of indices, may be even farther removed from actual mental workings than information states (which are refinements of propositions in that every information state determines a unique corresponding proposition but not vice versa). But information states are still propositions of sorts and *ccp*s are not. So once more: What role can *ccp*s play in an agent's mental life?

Here is our answer: The roles played by *ccp*s are indirect: *ccp*s contribute through the information states (and, derivatively, the propositions) they yield when they are applied to information states, and in particular to the information states determined by the content specifications on which they referentially depend. Just a couple of examples for now of what form the contributions that *ccp*s can make through such applications.

First, the *ccp* $[[K_{DES}]]_M$ determined by the desire constituent of (2.3) makes its contributions to the mental life of an agent A whose mental state can be described by (2.3). In combination with the content of the belief in (2.3) this *ccp* can account for some aspects of how A will reason in certain situations about ways of making her desire come true. The *ccp* will make its contribution to such bits of practical reasoning via the information state $[[K_{DES}]]_M([[K_{BEL}]]_M)$ that we get by applying it to the information state $[[K_{BEL}]]_M$ determined by the belief. It might be objected that this cannot be quite since $[[K_{DES}]]_M([[K_{BEL}]]_M)$ captures not only the content specified by the desire constituent but the content specified by the belief constituent as well. Isn't this just the kind of lumping of desire and belief of which we observed earlier that that is what we do not want, in contrast with the harmless lumping that is used in DRT to account cumulative interpretations of discourse? No, not quite. In the first place, we still have the belief content as a separate semantic object. And in the second place, and more importantly, when reasoning about the desire – for instance in trying to figure a way to get it satisfied – it will be on the assumption that the belief on which the desire depends is true. The aim of the reasoning is to find a way of turning the world from one in which the belief holds into one in which the desire holds as well. In other words: in going from a world satisfying the belief content to one that satisfies the conjunction of the belief content and the desire content that depends on it. This is just the transition from a world that satisfies $[[K_{BEL}]]_M$ into one that satisfies $[[K_{DES}]]_M([[K_{BEL}]]_M)$.

A similar story can be told about the intention constituent and the belief constituent of (4.4), the MSD for Ponce de Léon's state of mind before he went on the expedition to Florida. The information state obtained by applying the *ccp* determined by the intention to the information state determined by the belief gives an information state that combines the contents of the two. But practical reasoning in relation to the intention – e.g. in planning the Florida expedition – will be about realizing a world in which the Fount of Youth is found and – of course – exists.

Furthermore, speculations involving the desire in (2.3) or the intention in (4.4) need not include the beliefs on which these components depend in their respective MSDs. They can also be combined with alternative assumptions about what the world is like, provided only that those alternative assumptions include entities that can serve as values for the drefs via which they are referentially dependent. For instance, the agent of (2.3) could think about the possibility that what she is seeing in the middle of the road isn't made of gold, or perhaps not a coin at all. But of course, her speculations have to maintain that there is something there, if her desire will simply dissolve as not about anything. Likewise, Ponce de Léon may have alternative thoughts about the whereabouts of the Fount of Youth. Or he may wonder what the taste will be of its water and whether it will be transportable in normal vats. That is, the *ccp*s determined by desire or intention can be applied to other information states in such speculations than those specified by the MSD constituents on which they are dependent in the minds of their agents. In these reflections the *ccp*s are applied to information states other than those determined by the constituents on which they depend in the actual current mental states of their agents.

These are just two examples of how the *ccp*s determined by referentially dependent MSD constituents can make their contributions to mental dispositions and processes. There is no a priori reason why *ccp*s determined by referentially dependent constituents in other cases can make their contributions in this same way, however, and more work will be needed to develop a clearer and more complete picture of how referentially dependent MSD constituents interact with the constituents they depend on, as well as with constituents on which they do not depend. But this work too will have to be left for some other time. Instead we return to the complication mentioned at the outset of the present section, the dependence of the hope constituent of (4.8) on both the knowledge and the belief constituent.

The drefs involved in these dependencies are h and l. According to the general principle we followed in defining the *ccp* for the desire constituent of (2.3) the *ccp* for K_{HOPE} should therefore be defined for those information states SAT whose Argument Domain includes h and l[9]. For any such SAT and index $<w,t>$, $[[K_{HOPE}]]_M(SAT)(<w,t>)$ consists of those assignment functions g with domain $ArgD(SAT) \cup \emptyset$ for which there is an assignment $f \in SAT$ such that $f \subseteq g$ and g verifies K_{HOPE} at $<w,t>$ in M.

This description of the *ccp* for the hope constituent is hardly more than a repetition of what we said about the *ccp* determined by the belief constituent of (4.8). But the two cases are not quite the same. Different about the case before us now is that since the hope constituent of (4.8) is dependent on more than one other constituent, the information state to which it should be applicable, according to the MSD,is an amalgam of some sort of the denotations of different constituents of the MSD. The question is what sorts of amalgams these are.

For the case at hand, where the hope specification depends on both the knowledge and the belief constituents, the amalgamation is simple; it consists of no

[9]and is disjoint from the empty set, but that is trivial for this example

more than applying the *ccp* denoted by the belief specification to the information state denoted by the knowledge specification. This leads to the information state $[[K_{BEL}]]_M([[K_{KNOW}]]_M)$ as amalgamation result. The intended application of the *ccp* for the hope constituent is to this information state, which results in the information state $[[K_{HOPE}]]_M([[K_{BEL}]]_M([[K_{KNOW}]]_M))$.

But this is still a comparatively simple case of referential dependency. Much more complex dependency structures are possible.[10] In general a given content specification K may referentially depend on specifications K_1, ..., K_n, and any of these K_i may depend in its turn on specifications $K_{i.1}$, ..., $K_{i.n_i}$, and so on. But *not* ad infinitum. We assume, as a constraint on well-formed MSDs, that the relation that holds between content specifications K and K' iff K referentially depends on K' is *well-founded*, i.e. that any chain of the form K_1, ..., K_k, ... , where all the K_i are content specifications belonging to the same MSD and where K_i is referentially dependent on K_{i+1}, is a finite chain. (Note that among other things this excludes loops: If K_1, ..., K_k is such a descending chain, then it won't be the case that K_k referentially depends on K_1.)

From here on we assume that all MSDs are **well-founded** and also that they are **finite** (in the sense of being finite sets).

Let MSD be an MSD that is finite and well-founded in this sense, and K the content specification of any of its Propositional Attitudes. Let RD*(K) be the set of all specifications K' in MSD on which K directly or indirectly referentially depends. That is, $K' \in$ RD*(K) iff there is some finite chain $K = K_1$, ..., $K_k = K'$ in which K_i is referentially dependent on K_{i+1}. Then the elements of RD*(K) can be ranked as follows. The K' that are at the end of maximal descending chains starting with K are not referentially dependent on any other specifications in MSD. These get rank 0. If all K' in RD*(K) get rank 0, then we are done. If not, then there will be K' in RD*(K) that depend only on specifications with rank 0. These get rank 1. If this still does not exhaust RD*(K), then there will be K' in RD*(K) that only depend on

[10] This claim is based on some general definition of the that set of well-formed MSDs that will have to be part of any formal definition of the syntax of a DRS language of MSDRT that includes all the examples of MSDs given in this introduction. We haven't provided such a definition so far, and won't do so either at any later point, trusting that the readers will be familiar enough with such definitions of the well-formed expressions of formal languages, that they would be able to come up with such definitions if pressed, and that can make sense of the following discussion on the basis of their knowing how to go about such definitions.

specifications of rank 0 or 1. These get rank 2. And so on until RD*(K) has been exhausted.

If RD*(K) is non-empty and r is the highest rank assigned to the members of RD*(K), then K itself is assigned rank r+1. Otherwise, if RD*(K) is empty, then K is assigned rank 0. Since MSD is finite, every content specification of a Propositional Attitude in it is assigned some finite rank r.

This ranking of the content specifications in MSD allows us to recursively assign information states to those specifications. First an auxiliary notion. Let SAT_1, ..., SAT_n be information states in M. The *merge* MERGE(SAT_1, ..., SAT_n) of SAT_1, ..., SAT_n is defined as follows: (i) at least one of SAT_i is the necessarily false information state, the one such that for all $<w,t>$ $SAT_i(<w,t>) = \emptyset$. In that case MERGE(SAT_1, ..., SAT_n) also is the necessarily false information state. (ii) none of the SAT_i is the necessarily false information state. Let D = $Dom(f_1) \cup ... \cup Dom(f_n)$. For each i = 1,..,n let SAT'_i be defined by: $SAT'_i(<w,t>) = \{g$: $Dom(g) = \text{D}$ & $(\exists f)(f \in SAT_i(<w,t>)$ & $f \subseteq g)\}$. Then MERGE(SAT_1, ..., SAT_n) = $SAT'_1 \cap ... \cap SAT'_n$.

The *primary information state determined in* model M *by* a content specification K in MSD, $[[K]]^{PR}_M$, can now be defined as follows by recursion on rank:

(4.14)a. Suppose that K has rank 0. Then $[[K]]^{PR}_M = [[K]]_M$, the information state determined by K in M, as defined in (4.6) in Section 2. (Thus, when K determines an information state, then its primary information state is by definition just that information state itself. but this is a kind of degenerate case, and we won't normally use the term 'primary information state' in cases of this kind.)

b. Suppose that $[[K']]^{PR}_M$ has been defined for all K' of rank $\leq$ r and let K be of rank r+1. Let K_1, ..., K_n be the specifications on which K (directly) referentially depends. Then each K_i will have rank $\leq$ r and thus $[[K_i]]^{PR}_M$ is defined. So the merge $[[K_1]]^{PR}_M \oplus ... \oplus [[K_n]]^{PR}_M$ of these information states is defined as well. We refer to this information state as the *primary input state to* the *ccp* $[[K]]_M$. $[[K']]^{PR}_M$, the *primary information state determined by* K in M is then the result of applying $[[K]]_M$ to this primary input state:

$$[[K']]^{PR}_M = [[K]]_M([[K_1]]^{PR}_M \oplus ... \oplus [[K_n]]^{PR}_M)$$

For each member $<MOD,K>$ of an MSD MSD let ANC(K) (for 'the ancestors of K') be the set of all constituents of MSD on which K directly or indirectly depends. Then the MERGE of ANC(K) can be defined as above.

Definition (4.11) makes no explicit reference to the contributions that ERs can make to the *ccp*s denoted by content specifications. This too constitutes a choice made by the semantics we are proposing. When a *ccp* has occurrences of the distinguished dref x of an ER ER, then it will be the case that when it is applied to an information state that is about the entity represented by ER, then the information state resulting from the application will also be about that entity: *ccp*s as defined in (4.11) preserve aboutness. But these *ccp*s can also be applied to information states that are not about particular entities, and the information states resulting from those applications won't then be about particular entities either. An alternative definition for *ccp*s would have been possible in which the entities represented by the ERs whose distinguished drefs occur as arguments in K are made integral parts of the *ccp*, in such a way that the *ccp* is defined only for information states that are about these entities. But to do this would have complicated the definition above without any benefits to the further development of our semantics.

The discussion of referential dependence in this section has led to one firm constraint on MSDs: the referential dependence relation should be well-founded. But we saw in Section 2.2.1 of Ch. 2 that MSDs are also subject to the constraints imposed by the Attitudinal Hierarchy. We also noted there that we still know far less about the Attitudinal Hierarchy. than we would like snd that we should. The Attitudinal Hierarchy is one of the seriously underdeveloped modules of MSDRT. For our concerns in Chapter 4, however, this lack of firm and detailed knowledge about the Hierarchy does not matter. The definitions we have given for the denotations of MSD constituents are sound whether or not the MSDs they belong to are coherent in the sense we have explored in the present section. Thus as far as the model-theoretic semantics for MSDs goes, we can leave further investigations into MSD coherence until later, in the hope of gradually flling in the many gaps in our knowledge about the Hierarchy.

4.2.3 Plans

There is one further point we want to raise before concluding Section 2. It has to do with the ultimate motivation for this MSDRT Introduction as part of our project to provide a logical representation language for the theory of individual and joint intentions, planning and action.

Some of our intentions are simple; they specify a single action, often with the aim of thereby accomplishing a certain result. For instance, you intend to pick up a coin from the middle of the road with the aim of having it in your possession. But not all simple intentions are like this. Sometimes the intended action is an aim in itself, as when you intend to dance or to take a walk, just for the pleasure of doing so.

On the other hand, many intentions are complex in a way that neither of these two intentions are. Their aim is not a single action, but several actions, which have to be performed in some more or less fixed order. Often such complex intentions are referred to as 'plans'. Typically a plan consists of a temporally ordered set of intentions whose execution is to issue in the realization of a certain goal. And often the goal is there first and the plan – in the narrow sense of this word, in which a plan is separate from its goal – is then constructed as a way of getting to the goal. In general, plans can be of considerable complexity, involving clusters of actions that are required to jointly achieve the final goal or some sub-goal, and that can or must be performed simultaneously or can be performed in a number of different orders, and/or alternative courses of action, to be armed against situations in which one of these turns out to be impossible to execute. And there is also the dynamics of plan design, with more details getting added to the plan in the course of time, and even after the plan's execution has begun, in these cases where execution itself takes time and new possibilities and constraints are discovered along the way.

All this is for later. Here we only have a quick look at a comparatively simple type of plan, a *linear* plan, which consists of a series of actions that lead to the intended goal when they are executed one after the other, with each execution creating a state of affairs that enables execution of the next action. Good examples of such plans are those for getting from one spatial location to another. Suppose that our agent has to get from her home in city A to a certain address in city B and that she has formed to this end the plan informally described in (4.15).

(4.15) Buy a train ticket for the trip form A to B on the internet; go, at the right time, from home to the bus stop of the bus that gets me to the main railway station in A; wait for the bus to the station; when it comes, get on it; get out at the station; go to the platform where the train to C leaves; wait for the train for C to arrive; when the train to

> B arrives, get on the train to C, get out of the train at C; go to the platform were the connecting train to B leaves; wait for the train to B; get on this train, when it arrives; get out from the train at B; go out of the train station in B; find a taxi; get in and tell the driver to get me to the address where I have to be.

The details of this plan are a little arbitrary; there could have been more details, with what is described here as a single action broken up into more parts, e.g. more details about the online purchase of the ticket; or the actions between arriving at C and leaving from C could be bundled into 'change trains at C'; and so on. But that is the way with the plans we make: the actions that are less routine, or whose execution worries us more, we will try to plan in greater detail. And details may be added as time goes on and the time for execution approaches.

The goal of this plan is to be at the given address in B. That is a state of affairs, which is the result of the event implied by the final action of the plan, that of telling the driver where he should go. (It is presumed in this plan that this will have the effect of the driver going to the address you have given him and to deliver you there.) On the other hand, the last action, of telling the taxi driver in B where to go and the taxi trip that results from this, are possible only in the state that has just been reached through the one but last action – that of having found a taxi in B. And so it goes, all the way back to the purchase of the ticket.

There are two ways to think about the formal representation of linear plans like the one in (4.15). The first takes the form a single intention whose content is a kind of narrative structure, of a sequence of successive events, each with its own result state, and where the result state is a precondition for the next event happening. If the contents of the successive items of the plan are represented in the form of DRSs K_1,..., K_{16}, where for each K_{i+1} its Universe contains an event dref e_{i+1} and a state dref s_{i+1} and its Condition Set the Conditions '$e_{i+1} \subseteq s_i$' and '$e_{i+1} \supset\subset s_{i+1}$'.The first of these Conditions shows that K_{i+1} is referentially dependent on K_i. The content specification of the intention is the merge K_{tot} of all the K_i - a DRS of the kind discussed in (Partee 1984)[11] Note that in this merge DRS the local dependencies are taken care of through merge, in the familiar manner of DRT.

Another way to think of the plan in (4.15) is as a succession of intentions, one for each of its 16 clauses. Since all K_i other than K_1 are referentially

[11]For some more discussion of such structures see also (Kamp 2019).

dependent on their predecessor, they all denote *ccp*s that are not information states. But in view of our assumption that PRD is reflexive, this does not prevent us from representing the plan as the set of intentions $\{<INT,K_1>$, ..., $<INT,K_{16}>\}$, which could play the part of an MSD all on its own, or as one that is part of some larger MSD. Note that when the *ccp* denoted by K_2 is applied to the information state denoted by K_1, and then the *ccp* denoted by K_3 to the information state resulting from the first application and so on, then the final information state will be the one denoted by the specification K_{tot} of the single intention described in the last paragraph.

The point of discussing this example has been to show that among the referential dependency chains in MSDs there can be arbitrarily long ones, as well as the short ones considered up to now. A more detailed discussion of plan structure, with much more complex structures than the one described in (4.15) will be discussed in a sequel to the preset introduction.

4.3 Truth Conditions for Attitude Reports

In this section we finally come to the truth conditions for the Logical forms of attitude attributions. Our central task is to specify what it is for such a Logical Form to be a true description of the mental state of some agent a at some time t in some world w of a model M. Because of the structure of our Logical Forms this specification involves three successive semantic relations: (i) the correctness of the MSDs occurring in those Logical Forms qua descriptions of the relevant mental states in the model, (ii) the verification of the *Att*-Conditions containing those MSDs, and (iii) truth conditions for the DRSs containing those *Att*-Conditions.

Throughout this section we stick with our provisional assumption that the mental states of agents a at indices $<w,t>$ in models M have the form of MSDs, and thus are the same kinds of entities as their descriptions in the Logical Forms of mental state attributions. This assumption will be discussed and challenged only in Section 5.1.

We start with the correctness conditions for MSDs. As we will see, this is the main bottleneck in the truth definition for (the Logical Forms of) attitude attributions.

4.3.1 Correctness conditions for MSDs

The general gist of the correctness definition for MSDs that we will propose will have been more or less clear from the earliest parts of this Introduction to MSDRT: An MSD MSD is a correct description of a mental state MS if there is a part MS' of that state, consisting of some of the entity representations and/or some Propositional Attitudes, that is correctly described by MSD. But what can all this mean?

The first complication that we need to deal with is one of referential commensurability. Even if we assume, as we do for now, that the first component MS_1 of the pair MS_M of an MSDRT model M (see definition (4.1) in the opening section of Chapter 4) has the form of an MSD, there is no reason why the drefs occurring in it should bear any intrinsic relation to those in MSD. That is, part of the question which part of MS_1 can be chosen as the one for which MSD provides a description is how the drefs in MSD can be seen as 'translations' of the drefs in MS_1. Such a translation must therefore be an intrinsic part of the selection of MS_1'. For every MSD, whether constructed as part of the Logical Form for an attitude attribution or as someone's actual mental state, let DREF(MSD) be the set of all drefs that have one or more occurrences in it. Then taking MS_1' to be the part of MS_1 that is targeted by MSD presupposes a 'translation' function TRANS from DREF(MSD) into DREF(MS_1). We will assume that TRANS is an injection.[12]

The question, then, whether MSD is a correct description of MS reduces to whether there is a pair $<$TRANS,$MS'>$ such that (i) TRANS is an injection from DREF(MSD) into DREF(MS) and (ii) MS_1' is a subset of MS_1 that can regarded as the 'direct description target' of MSD. But what can it be for MS_1' to be the direct description target of MSD?

To facilitate the following discussion let MS' be the result of replacing in MS_1' all drefs α by TRANS$^{-1}(\alpha)$.

Here is our next decision point. But here too the decision is more or less dictated by the original intuition that led to MSDRT: MSDs are to capture not only the Modes and contents of individual attitudes of the attributees of attitude attributions but also the connections between different attitudes, with a special emphasis on the referential dependencies between them. This entails that MS' should not only contain propositional attitudes and entity representations for each of the Propositional Attitudes and ERs in MSD,

[12] That is, TRANS is one-to-one.

but that the dependency structure of MS' should also be like that of MSD. For instance, if MSD contains a belief with content description K_{BEL} and a desire with content description K_{DES} and K_{DES} depends referentially on K_{BEL} via the dref x, then MS' should also have a belief and a desire whose contents are connected via the dref TRANS$^{-1}(x)$.

Similar considerations lead to the constraint that when MSD contains a Propositional Attitude $<MOD,K>$ and the distinguished dref x of some ER belonging to MSD occurs in the Condition Set of K, then MS' must also have a corresponding attitude $<MOD',K'>$ and an entity representation with distinguished dref x that occurs in some of the Conditions of the Condition Set of K'.

To state these constraints formally, let TARGET be an injection of MSD into MS_1 with Range MS'_1. The correspondences referred to above are then correspondences between the constituents $<MOD,K>$ of MSD and the corresponding constituents TRANS^{-1}(TARGET($<MOD,K>$)) of MS'. We will denote those corresponding constituents of MS' also as 'TARGET($<MOD,K>$)', omitting the bit 'TRANS^{-1}'. In fact, we will henceforth simply talk about the functions TARGET without mentioning the translation functions TRANS they presuppose at all. But it should be kept in mind that such a translation function TRANS is always there in the background.

Our first use of this convention is in (4.16), which states the constraints that are imposed on the embedding functions TARGET (and their underlying translation functions TRANS).

(4.16)

Let us assume, then, that MS' and TRANS allow for a 'dependency-faithful' embedding function TARGET from MSD onto MS'_1: a bijection between MSD and MS'_1 with the following properties:

(i) TARGET preserves attitudinal Mode: beliefs are mapped onto beliefs (formally: a Propositional Attitude $<BEL,K>$ is mapped onto some Propositional Attitude $<BEL,K'>$), desires onto desires, entity representations onto entity representations etc.

(ii) TRAGET preserves content. Exactly what that comes to will be explained below.

(iii) TARGET preserves referential dependencies: if MSD contains Propositional Attitudes $<MOD_1,K_1>$ and $<MOD_2,K_2>$ and K_2 depends referentially on K_1 via the drefs x_1,..., x_n, then it will be the case that TARGET($<MOD_2,K_2>$)

referentially depends on $<MOD_1,K_1>$ via TRANS(x_1),..., TRANS(x_n).

(iv) TARGET preserves singular reference: if MSD contains a Propositional Attitude $<MOD,K>$ and an ER ER with distinguished dref x and x occurs as argument in the Condition Set of K, then TRANS(x) will be the distinguished dref of the ER TARGET(ER) and will occur in the Condition Set of the DRS K' of TARGET($<MOD,K>$).
Here an interim summary. The point we have reached in our definition of the correctness of MSDs as descriptions of agents' mental states is this: For the MSD MSD to be a correct description of the mental state MS, there must be a translation function TRANS from DREF(MSD) into DREF(MS), a part MS_1' of MS_1 and a function TARGET from MSD to MS_1' satisfying the conditions in (4.16) such that the constituents of MSD stand in the right relations to the constituents of MS_1'. More specifically, TARGET must preserve Mode and content as well as referential dependence between constituents.

The one notion in (4.16) that needs further explanation is preservation of content. This notion is relevant only for Propositional Attitude constituents.[13] There are three main problems we have to deal with in this connection. The first is that the right notion of content preservation depends on the attitudinal Mode of the Propositional Attitude. The second has to do with the fact that the denotations of content specifications are often *ccp*s that are not information states. And the third, and most difficult problem is that our correctness definition has to proceed by recursion, starting with content specifications that do not contain any occurrences of *Att*. The reason for this is that in specifying the content of a content specification in a model M we have to rely on the model-theoretic semantics for DRT (see Chapter 1, Section 3). This will enable us to formulate the verification conditions for *Att*-predications whose MSDs do not contain any occurrences of *Att*. Once this has been done, we have therewith also defined the denotations of DRSs that contain such *Att*-predications (though not yet for DRSs in which *Att*-predications may contain other *Att*-predications). The verification conditions for such 'second order' *Att*-predications become available at the next round; and so on, for any finite degree of embedding of *Att*-predications within each

[13] The only preservation requirement we adopt for ERs is that if $ER \in MSD$, then TARGET(ER) is also an ER. The question in what sense TARGET must preserve the denotations of ERs will be discussed later. It is a point of debate whether TARGET should also preserve all or part of the descriptive content of ERs ad/or some or all of its anchor set. We set these questions aside here, to be addressed as part of a more general discussion of alternatives to the semantics developed in Chapter 4, which is set aside in this document.

other.

We start with the question how content preservation for a Propositional Attitude constituent $<MOD,K>$ of MSD depends on the Mode Indicator MOD. The felevant distinction here is between two types of Mode Indicators, *Exhaustivity Requiring* (EXREQ) Indicators and *Non-Exhaustivity Requiring* (NON-EXREQ) Indicators. EXREQ Indicators are Indicators of those attitudinal Modes for which correctness requires that the report matches the content of the reported attitude in full: if TARGET($<MOD,K>$) = $<MOD,'K>$, then K must express the same content as K'. For NON-EXREQ Indicators this is not so. Here K is only required to express part of the content of K'. That is, correctness only requires that K' entails K.

From the perspective of the traditions in the philosophy, logic and linguistics of propositional attitudes it is the NON-EXREQ Indicator type that is the default case. For instance, BEL, DES and INT, the Indicators focused on in earlier work on MSDRT are NON-EXREQ, and so is the Indicator $KNOW$ from (2.35). Among the EXREQ Modes are doubt and wondering (whether something is or isn't true). Indicators for these Modes should be part of any self-respecting version of MSDRT too. But they haven't been so far. And a systematic treatment of them jointly with BEL, DES and INT and perhaps other NON-EXREQ Indicators has to our knowledge not yet been undertaken, either within the setting of MSDRT or elsewhere.

Before we go on with the formal aspects of MSD correctness, here a few informal considerations in support of the distinction bertween EXREQ and NON-EXREQ attitudinal Modes. Start with belief. An evident feature of belief attribution is that its content ascription need not exhaust the content of the belief it reports. An example is (4.17.a). This attribution seems fine even if part of the reported belief – that it was in the *South* of France – is not mentioned in the report. Note also that the attribution in (4.17.a) would be proper even when it is made on the strength of John's own saying that Mary went on holiday in the South of France (and not just that she went on holiday in France) and also that it would have been perfectly all right for John himself to have expressed his belief as 'I believe that Mary went on holiday in France', knowing well that that it was the South of France.

(4.17)a. Fact: John believes that Mary went on holiday in the South of France.

Attribution: *John believes that Mary went on holiday in France.*

b. Fact: Mary wants to marry a Swede who is tall, handsome, blonde

and suave.
Attribution: *Mary wants to marry a Swede.*

c. Fact: Michael intends to go to NYC by train.
Attribution: *Michael intends to go to NYC by train.*

This possibility to under-describe the content of a belief as the agent understands it herself in a report of that belief is a general feature of belief reports. And we find this not only with beliefs but also with desires, intentions and a number of further attitudinal Modes. (4.17.b) and (4.17.c) are further examples of this. Our conclusion: BEL, DES and INT are among the NON-EXREQ Mode Indicators.

But as noted above, not all attitudinal Modes are like this. Doubting, for one, isn't. If Fred has expressed his doubt that John and Mary went to the South of France together, it could be quite wrong for you to report that doubt by saying "Fred doubts that John went to the South of France" For it may well be that Fred has no doubt whatever that John went to the South of France. What Fred doubts is that it was together with Mary that he went there. So doubt, with Mode Indicator DBT, is an EXREQ Mode. The same is true for the attitude we call *wonder*, with Mode Indicator WON, the attitude that an agent has towards a propositional content p iff she is undecided whether p is to be considered true.[14]

The correctness account we propose captures the difference between the correctness of MSD constituents with NON-EXREQ and those with EXREQ Mode Indicators in terms of relations between the content specification K of the attribution and the content specification K' of its target: correctness of a NON-EXREQ attribution requires that K' entails K; correctness of an-EXREQ attribution requires that K' entails K and conversely that K entails K'. But what should be understood by 'entailment' in the setting of our semantics for MSDs?

Here we reach anther decision point for our proposal. But once again the decision follows in the wake of others that have already been made along the way: our semantics is an intensional one, which makes use of an intensional

[14]This is typically the attitude of someone who sincerely asks if p is true. As with speech act types other than asking a question, their sincere performance entails a certain propositional attitude on the part of the speaker towards the propositional content p of her utterance. This connection between attitudes and speech act types is one of the reasons why WON must be included in versions of MSDRT that are designed as foundations for a combined theory of propositional attitudes and speech act types.

ontology, which includes the assumption that propositions are sets of indices and that entailment between propositions is set-theoretical inclusion: p *entails* q iff $p \subseteq q$. This entailment relation, however, is not the one we need, since our semantics for MSDs is not to be stated in terms of propositions, but in terms of information states and *ccp*s, and for these types of entities we need different definitions of entailment.

We start with information states. The entailment relation we adopt for these is standard from the Dynamic Semantics literature (Groenendijk, Stokhof & Veltman 2004). It not only requires that when the information state SAT' entails the information state SAT, then the indices that occur in SAT' also occur in SAT (i.e. that the proposition derived from SAT' entails the proposition derived from SAT), but also that the Argument Domain of SAT', ArgD(SAT')[15], includes the Argument Domain ArgD(SAT) of SAT, and that for any index $<w,t>$ and any assignment f' in $SAT'(<w,t>)$ is an extension of an assignment in $SAT(<w,t>)$. The definition is given in (4.18).

(4.18)Let SAT and SAT' be information states in a model M. Then SAT' *entails* SAT iff

(i) ArgD(SAT) is a subset of ArgD(SAT');

(ii) for every index $<w,t>$ and every assignment f' in $SAT'(<w,t>)$ there is an assignment f in $SAT(<w,t>)$ such that $f \subseteq f'$.

N.B. It is not hard to see that when SAT' entails SAT according to (4.18), p is the proposition derived from SAT and p' the proposition derived from SAT', then $p' \subseteq p$ (i.e. that p' entails p in the familiar sense mentioned in the paragraph above (4.18)).

Exercise: Verify this!

For *ccp*s – more precisely: *ccp*s that cannot be identified with information states – we need yet another notion of entailment, suitable for semantic entities of this type. *ccp*s are functions from information states to information states, and a natural concept of 'entailment' in the sense of 'at least as much information' is a 'pointwise' one: *ccp* CCP' entails *ccp* CCP iff applying CCP' to a certain information state produces an information state that contains at least as much information as the one we get when we apply CCP to that information state. But this formulation won't quite do for the relation that we need. The reason is that the *ccp*s CCP and CCP' that need to be

[15]See (4.5) in Section of 7.2.

compared in the correctness definition for *ccp*-denoting terms of 'entailment' will be from a Logical Form MSD and a mental state MS, respectively, and the latter will as a rule involve SATs with larger Argument Domains than the former. Our entailment definition for *ccp*s will have to take this difference into account.

The way we propose to account for the difference is as follows. Suppose that the content specification K of a constituent $<MOD,K>$ of MSD denotes a *ccp* CCP of M, that TARGET($<MOD,K>$) = $<MOD',K'>$ and that K' denotes in M the *ccp* CCP'. Then for CCP to count as a correct description of CCP' the first condition that must be fulfilled is that if CCP is a *ccp* for $PRES$ and D and CCP' a *ccp* for $PRES'$ and D', then $PRES \subseteq PRES'$ and $D \subseteq D'$. But the central condition is that when the information state SAT is in the Domain of CCP and the information state SAT' is in the Domain of CCP' and SAT' entails SAT in the sense of Definition (4.18), then this entailment relation is preserved by applying CCP to SAT and CCP' to SAT'.

We also need a notion of equivalence between *ccp*s to define correctness for MSD constituents with EXREQ Mode Indicators. As for information states, equivalence between *ccp*s is a stronger relation than entailment. But since the Domains of *ccp*s can differ, *ccp* equivalence cannot be simply defined as mutual entailment. However, we can approximate such a definition by observing that when CCP' entails CCP in the sense just described and defined explicitly in (4.19.a), then any information state in the Domain of CCP' is also in the Domain of CCP. This will be our extra condition for equivalence between *ccp*s.

(4.19.a) gives the official formulation of our notion of entailment between *ccp*s and (4.19.b) that of entailment between *ccp*s.

(4.19)a. Let CCP and CCP' be *ccp*s of a model M. Then CCP' *entails* CCP iff:

(i) Given that CCP is a *ccp* for $PRES$ and D and CCP' a *ccp* for $PRES'$ and D', then $PRES \subseteq PRES'$ and $D \subseteq D'$, then $PRES \subseteq PRES'$ and $D \subseteq D'$

(ii) Suppose that SAT is an information state in the Domain of CCP and SAT' an information state in the Domain of CCP' and that SAT' entails SAT (in the sense of Definition (4.18)). Then this entailment relation is preserved by applying CCP to SAT

and CCP' to SAT': $CCP'(SAT')$ entails $CCP(SAT)$

b. Let M, CCP and CCP' be as above. The CCP is *equivalent to* CCP' iff (i) CCP' *entails* CCP; and (ii) for each information state SAT' in the Domain of CCP', $CCP(SAT')$ entails $CCP'(SAT')$.

Exercise In Section 2.1 we observed that information states can be identified with *ccp*s of a special kind. In view of that, (4.19) can be understood as defining entailment and equivalence for information states, as a special case. We already gave definitions of those notions for information states in (4.18). Show that the definitions in (4.19) and those in (4.18) are equivalent. [end exercise]

We are now ready to define what it is for an MSD MSD to be a correct description of a mental state MSD' of an agent a at an index $<w,t>$ in a model M, but only for those cases where MSD' and MSD' do not contain occurrences of *Att* (since only for such MSDs the denotations of their content specifications are defined). Recall that correctness of an MSD, via a translation function TRANS and an embedding function TARGET, is defined in terms of correctness of each of its Propositional Attitude constituents $<MOD,K>$. We require entailment of $[[K]]_M$ by $[[\text{TARGET}(K]]_M)$ for those constituents for which MOD is NON-EXREQ, and equivalence for the constituents for which MOD is EXREQ.

(4.20)a. Let $<MOD,K>$ be a constituent of MSD with MOD a NON-EXREQ MOD, and let $<MOD,K'> = \text{TARGET}(<MOD,K>)$ be the corresponding propositional attitude constituent of MS'. Let CCP be the *ccp* in M that is determined by $<MOD,K>$ and let CCP' be the *ccp* determined in M by $<MOD,K'>$. Then $<MOD,K>$ *is a correct description of* $<MOD,K'>$ iff CCP' entails CCP .

b. Let $<MOD,K>$ be a constituent of MSD with MOD an EXREQ MOD, and let $<MOD,K'> = \text{TARGET}(<MOD,K>)$ be the corresponding propositional attitude constituent of MS'. Let CCP be the *ccp* in M that is determined by $<MOD,K>$ and let CCP' be the *ccp* determined in M by $<MOD,K'>$. Then $<MOD,K>$ *is a correct description of* $<MOD,K'>$ iff CCP and CCP' are equivalent.

As it stands, these definitions only work so long as the content-specifying DRSs only contains drefs of types (i) and (iv) in the sense of Def. (4.2). When content specifications also contain drefs of types (ii) or (iii), then the correctness definition has to be made to depend on an assignment f to all drefs of these two types. And for any dref α of type (iii) – in other words, α is the distinguished dref of some ER ER in MSD – $f(\alpha)$ must be the entity represented by the ER in a's mental state at $<w,t>$ that corresponds to ER via TARGET: $f(\alpha) = \mathrm{REF}_M(\mathrm{TARGET}(ER))$. (That is: f should assign α the entity represented by the ER from the mental state of a at $<w,t>$ that corresponds via TARGET to the ER ER of MSD.) When no such entity exists, i.e. when the ER in a's mental state at $<w,t>$ that is targeted by ER according to TARGET does not properly represent, then the correctness definition is no longer applicable as it stands, and some alternative notion of correctness is required to account for attitude attributions that we judge to be correct. (We have encountered one such case so far in the uses of the empty name *Vulcan* in Section 6. We will return to the question of correctness of such attributions in Section 6 of Ch. 4.)

There are various consequences this correctness definition ought to have and verifying that these follow from the definitions is a way of convincing oneself that the definition captures what it should. Here is, for starters, a very simple consequence:

(C1) Suppose that MSD has Propositional Attitude constituents $<MOD,K>$ and $<MOD',K'>$, that K is self-contained and that K' referentially depends on K and on no other Propositional Attitude constituents in MSD. Let M be a model and let TRANS be a 1-1 function from the drefs occurring MSD into DREFS(MS_1) and TARGET a function from MSD into MS_1, where MS_1 is the first component of the mental state in M of agent a at $<w,t>$ (i.e. the first component of $MS_M(<<w,t>,a>) = <MS_1,REF>$). Suppose that MSD correctly describes MS_1 via TARGET and further that TARGET($<MOD,K>$) is the only constituent of MS_1 on which TARGET($<MOD',K'>$) depends in MS_1. Let $[[K]]_M$ and $[[K']]_M$ be the information state and the *ccp*, respectively, that K and K' determine in M and likewise for $[[K_a]]_M$ and $[[K'_a]]_M$, where TARGET($<MOD,K>$) = $<MOD,K_a>$ and TARGET($<MOD',K'>$) = $<MOD',K'_a>$.
Then the information state in M that results from applying the denotation of K' in M (viz. $[[K']]_M$) to the denotation of K (viz. $[[K]]_M$) is entailed by the result of applying $[[K'_a]]_M$ to $[[K_a]]_M$.

(C1) can be inferred directly from the definitions and its verification is straightforward. But note: one premise of (C1) is that 'TARGET($<MOD,K>$) is the only constituent of MS_1 on which TARGET($<MOD',K'>$) depends in MS_1'. This is an assumption that does not follow from the correctness of MSD as description of a's mental state at $<w,t>$. As we have defined it, correctness allows for the possibility that TARGET($<MOD',K'>$) depends on constituents of MS_1 that are not in the range of TARGET. In that case the *ccp* determined by TARGET($<MOD',K'>$) will in general not be defined for the information state (or *ccp*) determined by TARGET($<MOD,K>$). What will always be defined, however is the primary information state that TARGET($<MOD',K'>$) determines in M.

Note on the other hand that the assumption we made about MSD, viz. that K' only depends on K, means that $[[K']]_M([[K]]_M)$ is the primary information state determined by K' in M.

It is not hard to see that in this case the primary information state determined by K' is entailed by the primary information state determined by TARGET($<MOD',K'>$).

This claim holds generally:

CLAIM Suppose that MSD is a correct description via TRANS and TARGET of the mental state of a at $<w,t>$ in M. Then the primary information state determined by K in M is entailed by the primary state determined in M by TARGET($<MOD,K>$).

Giving an explicit proof of this claim is awkward, and gets pretty boring well before it is finished. But we recommend making a start with the proof, just to convince oneself that and why the claim is true.

Here is a summary of the correctness definition for MSDs that are parts of the Logical Forms for attitude attributions:

• Correctness of the MSD MSD as description of the mental state $MS_M(<<w,t>,a>)$ in a model M of an agent a in M at an index $<w,t>$ is in terms of correspondence function TARGET and a translation function for drefs presupposed by TARGET.

• Given such a function TARGET (and the function TRANS presupposed by it), correctness requires: (i) TARGET preserves the referential dependency structure of MSD: when $<MOD,K>$ referentially depends on $<MOD,K'$ via the drefs α_1, ..., α_k, then TARGET($<MOD,K>$) referentially depends on TARGET($<MOD,K'>$) via TRANS(α_1), .., TRANS(α_k). Likewise, when $<MOD,K>$ contains occurrences of the distinguished dref α of the ER ER, then TARGET($<MOD,K>$) will contain occurrences of the distinguished dref TRANS(α) of TARGET(ER).

(ii) The content denoted by the content specification K of the Propositional Attitude constituent $<MOD,K>$ of MSD stands in the right entailment relation to the content denoted by the content specifier K' of the constituent TARGET($<MOD,K>$) (which is equal to $<MOD,K'>$).

The entailment relation depends on MOD. When MOD is a NON-EXREQ Indicator, then the relation is entailment of the denotation of K by the denotation of K'; when MOD is an EXREQ Indicator, then the relation is equivalence between the two denotations.

The meaning of 'entailment' and 'equivalence' depends in the last analysis on whether the denotations are information states or *ccp*s that are not information states.

• The constraints above say nothing about entity representations. The role that entity representations play in MSD correctness can be appreciated only within the wider context of *Att*-predication verification, which we turn to in the next section.

What can be said at the present stage suggests that the ERs of MSD do not impose any real constraints on correctness: When ER belongs to MSD and TARGET(ER) = ER', then the correspondence between ER and ER' established by TARGET is to be taken as implying that ER, as the description of the entity representation ER in a's mental state that represents entity **d**, is thereby to be regarded as also representing **d**. Thus, when the denotation of a content specification K in MSD denotes an information state SAT and K contains occurrences of the distinguished dref of ER, SAT is to be construed as an information state about **d** too.

The impression that the ERs of MSD do not impose any real constraints on correctness will be qualified in the next section.

The most serious problem about the definitions in (4.20) is that they can be applied only when the content specifications occurring as part of the Propositional Attitude constituents of MSDs have denotations (as information states or as *ccp*s) in models for MSDRT. At the present stage of our proceedings the denotations of these content specifications are properly defined only if these content specifications are DRSs without occurrences of *Att*. For so far we are only on our way toward the verification of *Att*.predications and therewith for toward the denotations of the DRSs containing them. This means that the definitions in (4.20) can be used in defining the verification conditions only of those *Att*.predications only so long as the MSDs occurring in these *Att*.predications do not contain *Att*.predications in their turn.

However, once verification conditions for such *Att*.predications have been given, that will then also give us a definition of the denotations of DRSs containing such *Att*.predications. This will then give the denotations of content-specifying DRSs in MSDs that contain *Att*.predications of these still simple kinds. But that then will lead to the verification conditions for *Att*.predications that contain *Att*.predications as parts, and so on.

Unfortunately, as we will see in the next section such a recursion isn't as straightforward as this informal description we have just given of it may suggests.

This concludes the quest for a definition of correctness for MSDs occurring in the Logical Forms of attitude attributions. The correctness definition will be the basis for defining the verification conditions for *Att*-predications and of DRSs containing *Att*-predications in the next section.

4.3.2 Verification conditions for *Att*-Predications and DRSs containing them

Att-predications are DRS Conditions of the form 's: $Att(a,MSD,\text{Links})$'. To see more easily what is involved in the verification conditions of such DRS Conditions it may help to return to our first example of a DRS containing such a Condition, (2.5) of Section 1:

(2.5)

$$
\begin{array}{|c|}
\hline
s \quad t \quad a \quad c \\
\hline
t \prec n \quad t \subseteq s \quad \text{Agnes'}(a) \quad \text{coin'}(c) \\
s\text{: } \mathit{Att}\left(a, \left\{ \begin{array}{l}
\left\langle [\mathit{ENT}, x], \begin{array}{|c|}\hline x \quad s_2 \\ \hline n \subseteq s_2 \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array}, \left\{ \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\text{: see}(i,x) \\ \hline \end{array} \right\} \right\rangle \\
\left\langle \mathit{BEL}, \begin{array}{|c|}\hline s_3 \\ \hline s_3\text{: } \mathit{gold\ coin}(x) \\ \hline \end{array} \right\rangle \\
\left\langle \mathit{DES}, \begin{array}{|c|}\hline s_4 \\ \hline n \subseteq s_4 \\ s_4\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\
\left\langle \mathit{INT}, \begin{array}{|c|}\hline t \ e \\ \hline n < t \ e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \{< x, c >\} \right) \\
\hline
\end{array}
$$

The central part to the verification conditions for *Att*-predications is the correctness of *Att*'s MSD argument MSD. Let us focus on (2.5) as our example. So MSD is the MSD shown in (2.5). Two of *Att*'s other arguments, its state argument s and its agent argument a, indicate of which are the mental states that MSD is presented as describing. Suppose that M is a model for MS-DRT and that f is an assignment such that $f(s)$ is a state of M and $f(a)$ an agent of M. Relative to such an assignment f the relevant mental states are those of $f(a)$ at indices 'covered' by the state $f(s)$. Here we encounter a further complication, having to do with the existence of eventualities in IH models. It seems intuitively clear that events can not only have duration within a given world w – if the event e exists in w at all, then it will have a duration $dur(e)$ in w – but that they can also exist in more than one world. For instance, x's endeavor to go from place A to place B may have succeed in the actual world w_0, but may have failed in some alternative world w' that still coincided with w_0 at the time when x set off. Such cases are most naturally thought of as involving the same event e, which exists both in w_0 and in w' even after the point when these worlds part company, but which may have different durations and also end differently in them, with x having reached B in w_0, but not in w'. We want such events that exist in more than one world for the analysis of modal and counterfactual statements such as

"John did reach B but he might have failed.", "John would not have reached B, if the road had been blocked." and the like. Much the same goes for states, including mental states. For instance, we often say things like "Mary still believes that the elections were rigged, but might have stopped believing this some time ago." The semantics of such statements is best analyzed as involving states that exist not only at more than one time but also in more than one world. For the given example, the state of Mary's holding the described belief should exist in more than one world as well as at more than time.

For these and other reasons we must admit both events and states that their existence may cover indices that differ in their world as well as their time components.

For our present purposes, however, the possibility that eventualities can exist at more than one world is not essential. The MSDRT languages we have so far considered in this MSDRT Introduction do not contain any modal constructs, whose evaluations might result in shifts from one world to another. In other words, when a DRS K of such a language is evaluated in a model M at an index $<w,t>$, then in the course of this evaluation parts of K may be evaluated at indices $<w,t'>$ for times other than t. But the evaluation will never shift to indices $<w',t'>$ with $w' \neq w$.

As before we assume that the Logical Forms of sentences and discourses are evaluated at indices at which those sentences or discourses are uttered. For DRSs of the languages considered in this Introduction this means that when a Logical Form is given as a DRS K from one of these languages and K is evaluated at an index $<w,t>$ (at which an utterance of this sentence or discourse is made), then evaluation of any constituents of K that are part of this evaluation will be at indices $<w,t'>$ (with possibly different times t', but the same world w). And that means in its turn that for the purposes of such evaluations that start out at $<w,t>$ only that part of the model M is relevant that involves indices $<w,t'>$, with varying times t' but fixed world w. In other words, Evaluations of such DRSs that start at an index $<w,t>$ of a model M will be restricted to what we may call the *w-section* of M – that sub-model M_w of M that consists of the extensional models determined by indices of the form $<w,t'>$ for varying t'. Such models are single histories. They have been used extensively within DRT since its earliest beginnings and have been explored closely in the course of their use in DRT applications.[16]

[16]See (Kamp 1981*a*) (the original English version can be found in (Kamp 2017)), Ch. 5

The entities found in w-sections of an IH model M are the 'w-sections' of entities that may exist at indices $<w,t>$ for varying times *and* worlds, where the *w-section* of an entity $\mathbf{d}$ of M is that part $\mathbf{d}_M$ of $\mathbf{d}$ that exists at those indices of M that are of the form $<w,t'>$ for varying t'. In particular, the eventualities of M_w are entities ev with a duration $dur(ev)$, consisting of all and only those times t' such that ev exists at index $<w,t'>$. It is assumed (as it has been all along) that $dur(ev)$ is always some convex subset of the time structure of world w in M.

This holds in particular for the states occurring as first arguments of *Att*. And that presents us with a first problem for the verification conditions of *Att*-predications. Our models for MSDRT specify mental states of agents for indices $<w,t>$ where t is an *instant* of the time structure of M_w. But evaluation of an *Att*-predication s: $Att(a,MSD,$Links$)$ will involve an assignment to the dref s of a state $\mathbf{s}$ of M_w with a duration that is an *interval* of this time structure. What does this tell us about the role of MSD correctness in the verification conditions of 's: $Att(a,MSD,$Links$)$'? The answer to this question seems to us to be clear enough: For 's: $Att(a,MSD,$Links$)$' to be verified in M_w, by an assignment f that is defined for all the free drefs of this predication, and which will assign $\mathbf{s}$ to s and some agent $\mathbf{a}$ to the dref a, MSD must be a correct description of the mental state of $\mathbf{a}$ at $<w,t'>$ for each instant t' belonging to the interval $dur(\mathbf{s})$. Or put less formally: MSD must be a correct description of $\mathbf{a}$'s mental state throughout the duration of $\mathbf{s}$.

Note that while the formal implementation of this answer is straightforward, there is something odd about this answer. Intuitively the constituents of a mental state – both its entity representations and its propositional attitudes – have durations too. For entity representations this has been implied by much of what we have said about them, for instance that their anchor sets often grow over time, with each new recognition of the entity they represent. But our conception of propositional attitudes is plainly one according to which they typically have duration too. We ask ourselves for how long we have had a certain belief or when we abandoned a certain desire. And often enough we entertain such thoughts about the thoughts of others. A plausible modeling of mental states should do justice to the temporal extendedness of their existence. The way mental states are captured in the models for MSDRT that we are using right now fail to capture this aspect of them.

The failure of our present set-up to do justice to the temporal extendedness

of (Kamp & Reyle 1993) and all but the first chapter of (Kamp 2015).

of mental state constituents manifests itself both in its models and in its DRS languages for MSDRT. It is a serious failure and one that should be corrected. But presenting the correction that we want to propose involves a number of issues and formal modifications that are more or less orthogonal to the central concerns of the present section. So we postpone these to the next section, Section 4, and continue for now with the other issues that are raised by the verification conditions for *Att*-predications.

So far we have just been concerned with the role of MSD correctness for the verification of an *Att*-predication. But there is more to the correctness of the MSD argument of such a predication than we have been able to explain in Section 3.1 of this Chapter. What is still missing is the role of the Links argument of *Att*.

Suppose that $<v,v'>$ is a member of Links in 's: *Att*(a,MSD,Links)'. Then that means that the entity **d** represented by the entity representation ER in MSD of which v is the distinguished dref has to be the same as the entity **v'** assigned to v' in the course of evaluating 's: *Att*(a,MSD,Links)'. This means that if the verification of the DRS K that contains 's: *Att*(a,MSD,Links)' is to be successful, then **d** must have all the properties that K attributes to **v'** through DRS Conditions other than those in MSD. The entity represented by the ER TARGET(ER) in the mental state of the agent **a** must have the properties that K asserts of v'. It is in this way that the correspondence between ERs in the describing MSD MSD and the mental state of **a** described by it makes a non-vacuous contribution to *Att*-predication verification.

For illustration consider the one link $<x,c>$ of the Links argument in (2.5). Verification of the *Att*-predication of (2.5) at some index $<w,t>$ of M_w will involve an assignment f that assigns entities in M_w for all the drefs that have free occurrences in this *Att*-predication.These are the drefs s, a and c. Suppose as above that $f(s) = \mathbf{s}$, $f(a) = \mathbf{a}$ and futher that $f(c) = \mathbf{c}$. Then verification of the *Att*-predication by f at $<w,t>$ in M_w will require functions TRANS and TARGET as explained in Section 3.1, and these have to be such that **c** is the entity **d** represented by TARGET(ER), (with ER the Entity Representation of (2.5)). On the other hand, the DRS (2.5) asserts of **c** that it is a coin. So verification of this DRS at $<w,t>$ in M_w by an assignment that includes f will impose this property on the entity **d** represented by TARGET(ER). Intuitively, the speaker of attribution to which (2.5) has been assigned as Logical Form makes a claim in this way about the entity **d** of which she assumes that the attributee a has an entity representation for it.

We are now nearly ready to state the verification conditions of the present section for *Att*-Conditions. Suppose that 's: *Att*(a,MSD,Links)' is this Condition and that it is evaluated at the index $<w,t>$ of M_w and relative to an assignment f to all the drefs with free occurrences in 's: *Att*(a,MSD,Links)'. Let us recall in this connection that some of these free dref occurrences can be in content specifications belonging to MSD. (These are the free dref occurrences of type (ii) in the classification (4.2). All in all the free dref occurrences in 's: *Att*(a,MSD,Links)' are of the following possible types:

(4.21)(a) the state dref s;

(b) the dref a for the attributee;

(c) drefs with occurrences in MSD that are not bound within MSD (drefs of type (ii) according to the classification scheme in (4.2));

(d) drefs occurring as second members of pairs in Links.

We have dealt with each of these four categories at this point. Let f be an assignment to each of these drefs. For any dref α of type (c) which occurs in the content specification K' of a constituent $<MOD,K'>$ of MSD, the denotation in M_w of K' is about $f(\alpha)$ in the sense explained in Section 3.1. Furthermore, if $<v,v'>$ belongs to Links, then the entity represented by TARGET(ER) where ER is the ER in MSD with distinguished dref v, must be identical with $f(v')$. And finally this must hold for all t' in the interval $dur(s)$.

The one remaining part in this informal description of the verification conditions of 's: *Att*(a,MSD,Links)' are the functions TRANS and TARGET. In Section 3.1 these were introduced in the context of correctness of MSDs as descriptions of mental states of agents at indices $<w,t>$, where t is an instant of time. But we have just seen that the verification of 's: *Att*(a,MSD,Links)' by f requires of correctness of MSD of the mental state of $f(a)$ throughout $f(s)$. The intuition that 's: *Att*(a,MSD,Links) expresses that a is in a mental state described by MSD *throughout* the duration of $f(s)$ implies that the sense in which MSD can be seen as describing this mental state, that $f(a)$ is in throughout $dur(f(s))$, remains constant throughout this period of time. This suggests that the functions TRANS and TARGET should remain constant throughout $dur(f(s))$. Adopting this last decision we come to the following statement of what it is for f to verify 's: *Att*(a,MSD,Links)' at t in M_w.

(4.22)(Verification of *Att*-Conditions (preliminary))

Let M_w be the w-section of a model M for MSDRT, s: *Att*(a,MSD,Links)' an *Att*-Condition, f an assignment of entities from M_w to the drefs with free occurrences in 's: *Att*(a,MSD,Links)', $<w,t>$ an index of M. Then f *verifies* s: *Att*(a,MSD,Links) *at* $<w,t>$ *in* M_w *at* $<w,t>$ iff the following holds:

There are (a) a function TRANS defined on the set of the drefs occurring in MSD the Range of which is included in the drefs of each of the mental states $MS_{M_w}(f(a)(<w,t'>)$ for t' in $dur(f(a))$ and

(b) a function TARGET defined on MSD such that its Range is included in each of the sets $MS_{1,T'}$ of the first components of the mental states $MS_{M_w}(f(a)(<w,t'>)$ for t' in $dur(f(a))$ such that for all t' in $dur(f(a))$:

(i) MSD is a correct description of the subset TARGET(MSD) of $MS_{1,T'}$, and

(ii) for each link $<v,v'>$ in Links $f(v') =$ REF(TARGET(v)).

Towards the end of Section 3.1 we observed that *Att*-Conditions are the only DRS Conditions that distinguish the DRSs of DRS languages of MSDRT from those of the underlying DRS languages of DRT. So once we have the verification conditions for *Att*-Conditions, we also have the verification conditions of the DRSs of MSDRT, and therewith also their denotations in MSDRT models. We also noted, however, that a definition like (4.22) must be applied recursively: Initially (4.22) is applicable only for *Att*-Conditions with MSDs whose content specifications are DRSs from the underlying DRS language of DRT, for which the denotations are determined by the independently given model theory for DRT. These content specifications do not contain any *Att*-Conditions. We suggested in Section 3.1 that in this way we can get denotations for content specifications that contain *Att*-Conditions which do not contain *Att*-Conditions in their turn. So Definition (4.22) can now be applied again, and to *Att*-Conditions whose MSDs may include content specifications that contain such simple *Att*-Conditions. This gives us the denotations for DRSs that contain *Att*-Conditions which contain *Att*-Conditions in their turn – 'second order *Att*-Conditions' – and so on, covering *Att*-Conditions of any finite degree of embedding depth and for the DRSs containing those. By repeating this procedure over and over we get a model theory for the DRS languages of MSDRT.

It should be made clear, however, that this construction only works as described, when at each stage of it the *Att*-Conditions to which (4.22) applies at that stage get *complete* verification conditions for each index $<w,t>$ of M and each assignment f of entities to the free drefs of any such *Att*-Condition the definition should tell us whether f verifies the *Att*-Condition at $<w,t>$ in M or not. At first sight it might seem that the definition does this. Let 's: *Att*(a,MSD,Links)' be such *Att*-Condition, and let M, $<w,t>$ M, and f be as described. Then either there are TRANS and TARGET as specified in (4.22) so that the conditions (i) and (ii) of (4.22) are satisfied, or there is no such pair of functions TRANS and TARGET. In the first case we have verification by f, in the second case we do not.

Unfortunately, as we hinted in Section 3.1, there is a snag here. It is this. We have been assuming that MSDs not only serve as constituents of *Att*-predications but also that the actual mental states of agents in our models are MSDs. The correctness definition for MSDs that occur as constituents of *Att*-predications compares the denotations in M of the content specifications of such an MSD MSD with the denotations in M of the content specifications from the mental states that MSD purports to describe. But that presupposes that these denotations are well-defined. And that is something that we cannot assume at the outset. For in general the content specifications of propositional attitude constituents of MSDs may contain *Att*-Conditions, viz. in MSDs that are part of the Logical Forms of iterated attitude attributions. And the denotations of such content specifications are not defined yet, since these denotations presuppose verification conditions for the *Att*-Conditions they contain. That is precisely what Definition (4.22) is to supply us with.

It was to cope with this difficulty that we proposed recursive iteration of (4.22). And that is fine so long as only the denotations of content specifications are concerned that occur in the Logical Forms of attitude attributions. Here it doesn't matter whether we get the verifications conditions for *Att*-Conditions in one fell swoop or through a potentially infinite number of iterations of some given procedure. But for the MSDs that *are* the actual mental states of agents in models the matter is different. The mental state of an agent at a given index $<w,t>$ of a model M just is what it is. If it includes attitude attributions that have attitude attributions as their content (or as part of their content), then that is a fact, and we need to respect it as that.

The description of such a mental state, however, may be partial, as it typically is for attitude attributing propositional attitudes whose Mode is NON-EXREQ, and such a description need not make any reference to the attitude attributing part of the state it purports to describe. It is this configuration that causes trouble for our strategy to obtain complete verification conditions for all *Att*-Conditions. In fact, the difficulty shows up right at the beginning, in the very first application of (4.22). This first application of (4.22) should give us complete verification conditions for MSDs that have no occurrences of *Att*, but it fails to do that because of the cases where the mental state that such an MSD purports to describe is itself an MSD that does contain occurrences of *Att*.

Is there a way out of this? Here is a suggestion. It may carry the smell of the type of stock market transaction that is known as the buying of futures. We assume that the entailment and equivalence relations between DRSs are given that we need in applications of (4.22) and assume that when we employ those relations in applications of (4.22), then those applications will confirm those relations.

Here is how this should work. We assume that there are relations $\models$ and $\approx$ between DRSs from the given DRS language for MSDRT, which hold between two DRSs K and K' only when K' contains *Att*-Conditions of greater depth than K.[17] We do not fix these relations any further in advance. But they should be such that when they are used in checking correctness as part of the evaluation of an *Att*-Condition, and thereby lead to complete verification conditions for this *Att*-Condition, and thus to well-defined denotations for content specifying DRSs that contain this *Att*-Condition, then applying the correctness definition to MSDs containing such content specifications should confirm these relations.

This is rather a mouthful, so let us unpack it a bit more slowly and stepwise. Suppose that 's: *Att*(a,MSD,Links)' is an *Att*-Condition, for which MSD is an MSD whose content specifications do not contain *Att*. We want to determine whether an assignment f defined for s, a, the second members of the pairs in Links and the free drefs of MSD (those with occurrences within MSD that are not bound anywhere within MSD) verifies this *Att*-Condition in M at $<w,t>$. When $f(a)$ is not an agent with a mental state at $<w,t'>$ in M for every t' $dur_w(f(s))$, then we are done: f does not verify 's: *Att*(a,MSD,Links)' at $<w,t>$ in M. When $f(a)$ is an agent with a mental state $<MS_{w,t',1},REF_{w,t'}>$ ($= MS_M(<<w,t'>,f(a)>)$) at $<w,t>$ in M for every t' $dur_w(f(s))$, and none of the content specifications of $MS_{w,t',1}$ contain

[17] The *depth* of an *Att*-Condition is defined in the obvious way. (i) An *Att*-Condition 's: *Att*(a,MSD,Links)' in which MSD doesn't contain any *Att*-Conditions is of depth 1. (ii) When the maximum depth of the *Att*-Conditions in MSD is n then 's: *Att*(a,MSD,Links)' is of depth n+1.

Att, then Definition (4.22) decides whether f verifies 's: *Att*(a,MSD,Links)' at $<w,t>$ in M. And when one or more of the content specifications of $MS_{w,t',1}$ do contain *Att*, then the verdict is reached using $\models$ and/or $\approx$.

In this way we arrive at complete verification conditions for *Att*-Conditions 's: *Att*(a,MSD,Links)' in which MSD does not contain occurrences of *Att* and with that denotations for DRSs that contain such *Att*-Conditions. Consider now a case where K is the content specification of a constituent $<MOD,K>$ of some MSD MSD, that K does not contain *Att*, that MSD is presented as the description of a mental state $<MS_1,REF_1>$ of some agent **a** of M at an index $<w,t'>$ and that MS_1 contain content specifying DRSs with *Att*-Conditions of depth 1, but no content specifying DRSs with *Att*-Conditions of depth > 1. Assume that $<MOD,K>$ belongs to MSD and that, for some suitable choice of functions TRANS and TARGET, TARGET($<MOD,K>$) = $<MOD,K'>$, where K' is a DRS containing one or more *Att*-Conditions of depth 1. Then the denotation of K' is now determined, via the full verification conditions for the *Att*-predication or *Att*-predications it contains and with that the question whether the denotation of K is entailed by it or is equivalent to it in terms of the Definitions ??? and ???. Whether either of these semantic relations hold between K and K' should match the verdicts provided by $\models$ and $\approx$. For instance, K' should entail K according to Definitions ??? and ??? iff $K' \models K$. If $\models$ and $\approx$ have been chosen in the right way, then these bi-conditionals should hold for each where an *Att*-free MSD is proffered as the description of a mental state that contains *Att*-Conditions of depth 1.

Suppose that $\models$ and $\approx$ have been chosen so that the bi-conditionals of the last paragraph all hold. Then this successfully completes our first round application of (4.22) and we are ready for round 2. This second round confronts us with the next installment of the difficulty encountered in round 1: The difficulty now arises in connection with the correctness of MSDs whose content specifications contain *Att*-Conditions is of at most depth 1 – i.e. either no *Att*-Conditions at all or *Att*-Conditions of depth 1 but not of any higher depth – as descriptions of mental state MSDs that contain content specifications with *Att*-Conditions of depth 2, but no *Att*-Conditions of depth > 2. Once again verification of the correctness of the describing MSD may require the use of $\models$ and/or $\approx$, but when that has been done, we have an independent way of assessing entailment and equivalence by using Definitions ??? and ???, and these assessments should be consistent with $\models$ and $\approx$ for the DRS pairs in question. A similar situations arises with each new application round, and if all goes well, then after infinitely many rounds the extensions of the relations $\models$ and $\approx$ will have been reconstructed by the verification conditions for *Att*-Conditions of arbitrary depth.

But is it reasonable to expect that relations $\models$ and $\approx$ exist? To prove this would require many assumptions about the DRS languages in question and perhaps also about the models for those languages. This is a matter that we have not yet found time to think our way through and we cannot even give a meaningful hint to someone who may want to devote thought to it. On purely intuitive grounds, however, we feel that there ought to be such relations: There ought to be well-defined relations of entailment and equivalence between content specifying DRSs whether or not attitude attributions are, or are part of, what it is they specify.

But we are also well aware of the limited value of intuitive speculation about matters involving formalization. More work is still needed here, probably a great dal of it.

(4.23) is a revised version of (4.22) which builds the existence of suitable relations $\models$ and $\approx$ into the definition. (This revised definition is mightily unpleasant to read and not recommended. Everything that distinguishes (4.23) from the earlier (4.22) has been discussed in detail above; so (4.23) is included just 'for the record'. We will return to the problems connected with (4.22) and (4.23) in Section 5.[18]

This concludes the topic of the present section: the verification conditions of *Att*-Conditions on the assumption that mental states – and not just their descriptions – are MSDs. But we are not yet done with the model theory of MSDRT. There remain three further topics. First, as mentioned above,

[18]One aspect of MSDRT's DRS languages is that *Att*-Conditions can occur in all sorts of positions. In the examples we have displayed in what has been going on so far, *Att*-Conditions always are among the Conditions of the Condition Set of the main DRS, but that is just because a primary function of such conditions is to provide the core of the Logical Forms for non-embedded attitude attributions, which are typically uttered in the form of assertions or as assertion-like speech acts. But even with the usual repertoire of complex Conditions for the DRS languages of DRT on which languages for MSDRT are built, the expressive power of those MSDRT languages is vastly greater what appears directly from the examples we have shown. How much of this expressive power is relevant to the Logical Forms of utterances that people will be prepared and able to make is one matter that requires closer attention than we (or, we believe, anybody else) has given it. Another question is whether the model theory developed in the present section retains its plausibility for DRSs in which *Att*-Conditions occur in logically embedded positions, such as in the scope of negation or in the antecedents or consequents of conditionals.

there is the matter of the temporal and modal continuity of the constituents of mental states, whether Entity Representations or Propositional Attitudes. That there is such continuity for entity representations is implied by the various things we have said about their reuse at times after they are first introduced, in particular when the represented entity is encountered again and recognized. For Propositional Attitude constituents temporal persistence seems hardly less obvious – for instance, how else could we make sense of what it is for someone to stick to their convictions? The way in which we have set up our models for MSDRT, which specify the mental states of agents at temporal instants and provide for no foolproof means of determining whether a constituent of the agent's mental state at one instant of time is or is not the same as a certain constituent of her mental state at some other instant, fails to do justice of the cross-temporal (and cross-modal) identity of propositional attitudes and entity representations. In the next section we address this problem and develop a way to fix it.

The second topic is the ontology of mental states. So far we have been assuming that the mental states of agents are made of the same stuff as the mental state descriptions of MSDRT – like their descriptions they just are MSDs. Many philosophers of a metaphysical bend (as well as some that would not like to see themselves described this way) may find this decision unpalatable to the point of jeopardizing the whole enterprise of MSDRT. In Section 5 we discuss some possible alternatives to the blatantly representational ontology of mental states as MSDs. This will also throw a new light on the vexing problem that we found in working out the verification conditions for *Att*-Conditions along the lines we followed above.

Section 6, finally, addresses an issue that is essential to MSDRT's treatment of attitude attributions that impute non-referring ERs to the attributee. In Section 6 we presented a proposal for dealing with empty names such as *Vulcan*. In Section 3 of Ch. 3 we proposed to capture intuitions that certain attitude attributions involving non-denoting names are true and other such attributions are false in the form of certain Logical Forms for those attributions. We did not say anything in Section 3 of Ch. 3 about a possible model-theoretic semantics for those Logical Forms – we couldn't have since the model theory for MSDRT was still to come at that point. But now that we have presented such a model theory in the successive parts of Chapter 4, the model theory for the Logical Forms of Section 6 can no longer be ignored in good faith. It is clear that the model theory we have developed in this and the past sections won't do for those Logical Forms. (Our present model theory doesn't assign any truth conditions to those Logical Forms. Section

6 will address the question what should replace it.)

4.4 Mental State Constituents as Temporal Continuants.

When discussing the satisfaction conditions for *Att*-predications in the last section we had our first taste of a problem that has been in the background all along, viz. that the propositional attitudes we entertain and the entity representations in our entity libraries have temporal extension: they come into being at some time, exist for some time after that and come to an end eventually, because they have done their job and are discarded or because they die with the agent to whose mental states they belonged or because at some time in the course of the agent's life time they have simply been forgotten. We first drew attention to the temporal extension of mental state constituents when discussing the verification of *Att*-Conditions in the last section. But we could have done so much earlier, for instance when introducing the communication-theoretic approach to semantics in Section 2.4. There we described the effect of the speaker's utterance (2.45), repeated below, on the recipient as the adoption by him of a *new* belief that the speaker read such a story the night before, which seems reasonable enough.

(2.45) I read a short story by Gogol.

But suppose the speaker had continued with "It is called 'The Nose'." Would this have led the recipient to form a new belief (referentially dependent on the one he already had) that the story spoken of in (2.45), and represented in his first belief, was entitled 'The Nose'? Or should the information he gets through this last sentence be treated as an update of the belief he had formed already? We find it hard to arrive at a clear judgment.

In principle the issue arises whenever MSDs are presented for the mental states of the same agent at different indices: Assume that $<MOD,K>$ and $<MOD',K'>$ are constituents of MSDs MSD and MSD' and that these are, or are descriptions of, a given agent A's mental states at two distinct indices $<w,t>$ and $<w',t'>$. (It may be convenient to think of these as just involving two different times t and t' in the same world w; that is, that $w' = w$.) Is there a way of telling if $<MOD',K'>$ is the same attitude or entity representation as $<MOD,K>$?

Cross-temporal identity of propositional attitudes raises a number of questions we need to distinguish. First there is the distinction between MSDs as constituents of the Logical Forms of attitude attributions and – in keeping with our current assumptions – MSDs as the mental states of agents. Let us first have a look at cross-temporal identity for constituents of MSDs in this second capacity. Suppose that $<MOD,K>$ and $<MOD',K'>$ are constituents of the mental state of A in a model M at the indices $<w,t>$ and $<w,t'>$. Is there a way of telling whether they are (manifestations of) the same propositional attitude or different ones?

To answer this question we need cross-temporal identity criteria for propositional attitudes. What could these be? A first shot might be that $<MOD,K>$ and $<MOD',K'>$ are manifestations of the same attitude iff they coincide both in their attitudinal Modes and in their content specifications. But what should that come to? As far as Mode is concerned the answer may seem simple: Each given DRS language for MSDRT specifies a set of Mode Indicators $MOD_1,..,MOD_n$. In such a language $<MOD,K>$ and $<MOD',K'>$ have identical mode iff $MOD = MOD_i$, $MOD' = MOD_j$ and i = j. But identity of content specification seems less straightforward. Is this supposed to be a matter of the *form* of the content-specifying DRSs or of their semantics (i.e. their denotations in the given model), or of something in between? This is a loaded question, and it has been that in particular in the context of DRT.[19] For the present discussion let us focus on the weakest of these possibilities: intensional identity, according to which the content specifications K and K' are identical (relative to a model M) if they have the same denotations (in M).

Even this comparatively weak criterion of content identity is arguably too strong for cross-temporal identity of proposition attitudes. One hint of this is the observation above that when the recipient who has already formed the belief that the speaker has read a short story by Gogol on the basis of her saying "I read a short story by Gogol." then gets the additional information that this story was 'The Nose' . Does this cause him to form a new belief or to update the old one? The second possibility may be no more than that. But even if it is just a possibility, it is arguably not one that we should exclude by the fiat of our choice of cross-temporal identity criteria.

[19]See for instance (Asher 1986), (Asher 1993), where the proposal is made to identify content for psychological purposes wth certain equivalence classes of logically equivalent DRSs.

The question what is to count as a single belief and what as two or more may often be hard to decide; and perhaps the decision should depend on how the agent herself handles her doxastic state at any one time. For certain purposes it may make sense to lump two beliefs together into a single one, e.g. as a single premise in reasoning, while for others one may want to keep them apart, for instance because one has a higher degree of confidence in the one than in the other. If this is true, then the cross-temporal identity of beliefs involves much less stability than the term 'cross-temporal identity' suggests. And in any case, as necessary and sufficient identity criteria for propositional attitudes as continuants, mode identity + intensional content identity won't do in general.[20]

With other attitudinal modes than belief our intuitions about when an attitude changes without losing its identity and when they are replaced by different ones may be guided by somewhat different considerations. But here too the combination of Indicator identity and intensional content identity seem to be too strong. Consider desires. One of the classical examples from the propositional attitude literature to show that desire is a NON-EXREQ Mode (Section 3.1) is the sentence "Mary wants to marry a Swede." This sentence can be used to describe Mary's want without implying that she would marry just ay swede and is acceptable as a correct description even when she has some further constraints in mind – that he should be handsome, considerate and reasonably well-off, as well as being blond perhaps and fairly tall; or whatever.[21] In this same spirit we may consider what could

[20]The criterion of Mode identity is not without problems either. The criterion may be fine for languages with a very small Indicator repertory, consisting, say, just of BEL, DES and INT. But with larger repertories the criterion may lose plausibility. Suppose for instance a repertory with a family of doxastic Indicators BEL_1,..,BEL_r which differ from each other in doxastic strength. Is it plausible that Mode Indicator identity is a necessary criterion for cross-temporal identity? Probably not. An agent may hold, over some given period of time, a belief with a given content but she may waver in the course of that period in how much confidence she has that her belief is true. So at different times t and t' from the period the beliefs she has at those times may differ in that it is $<BEL_i,K>$ at t and $<BEL_j,K>$ at t', with $i \neq j$. But should this force us to say that she had a *different* belief at t' from the one she had at t? That doesn't reflect the way we speak or think about such cases.

[21]There are also the uses of this sentence in which the phrase *a Swede* is used 'specifically' (in the terminology favored within Linguistics) – that use where some specific person that Mary has in mind is the one she wants to marry, and where this person is a Swede (which may but need not be one of her reasons for wanting to marry him; but she need not even know that he is a Swede). Such specific uses of indefinite DPs in desire attributions are treated in MSDRT as involving entity representations for the entities that justify the use of those indefinites, which sets them apart from the attitude reports discussed right here.

happen to Mary's marriage-related desires over time. She may at one point decide she wants to marry a Swede, but then, as time goes on, her desire may become more specific, imposing more and more conditions on what this Swede should be like. Does this mean that her desire is replaced by a different one each time a further constraint is added to her list? Perhaps that is the natural assessment in some cases, where the new constraint is a big one – adding it to the list changes the character of the desire, and the possibilities of its fulfillment, so drastically that it ought to be counted a new desire. But when the new constraints are more moderate, or even slight, describing Mary as having changed her mind, with a new idea of whom she wants to marry rather than having elaborated the desire she had, might seem arbitrary or outright wrong. Here too the cross-temporal identity criteria we adopt should not force us to have to say that even minor changes in content result in terminating the attitude.

A particularly strong case against the identity criteria we are discussing are intentions. One of the central points of Bratman's seminal (Bratman 1987) is that many of our intentional actions are the final result of intentions that we formed well before the time we execute them, and that in the course of their existence they may be subject to changes in the specification of what is intended. Often these changes take the form of expanding an initial intention, which specifies no more than a goal, wth a plan for how to reach this goal. And plans are quite often, perhaps even typically, reached in stages: they get modified – changed in the light of newly available facts, or gradually elaborated in greater detail – stepwise, but it seems natural with most such changes that it is still the same plan. It may be debatable whether plans are their own attitudinal Mode. But if they are, then certainly they would by a Mode for which the contemplated criteria would be too strong. According to our intuitions – and Bratman's too, if are not seriously misreading him – it is of the essence of plans that they can be modified over time and yet that what results from the change may still be the same plan as the one before the change.

Note that these examples all involve NON-EXREQ Modes. For attitudes with EXREQ Modes the identity criteria under discussion seem to us to be much less problematic. But even if these criteria can be considered adequate for EXREQ Modes, the fact that they fail for NON-EXREQ Modes remains;

For discussion see (Kamp & Bende-Farkas 2019). But the question may nonetheless be asked whether specific uses of indefinites in attitude attributions constitute the extreme of an otherwise continuous range stretching from the very unspecific to the more specific.

and since NON-EXREQ modes are central to any plausible conception of attitudinal states, the objection that the criteria are inadequate in relation to them is decisive. And note also: The intensional notion of content identity that is part of these criteria is a comparatively weak one. If this notion imposes constraints that are too strong, then the same will be true for any of the content identity criteria we mentioned above as possible alternatives.

Let us turn now to constituents of MSDs that are part of the Logical Forms of attitude attributions. Here the inadequacy of the identity criteria under discussion is even more blatant. Descriptions of NON-EXREQ attitudes, we noted, do not have to be exhaustive with regard to their content specifications. (That is what 'NON-EXREQ' means.) Let $<MOD,K>$ and $<MOD',K'>$ be constituents of MSDs MSD and MSD' that are part of the Logical Form of some complex attitude attribution, in which MSD is argument to an *Att*-predication involving an attributee a and a state holding at t in w of some model M and MSD' argument to an *Att*-predication involving a and a state holding at t' in w; and assume that MOD is a NON-EXREQ Mode and that $MOD = MOD'$. What can we say on the basis of a comparison between $<MOD,K>$ and $<MOD',K'>$ about the question whether they attribute the same attitude to a at t and t' or different ones? That the criteria considered in the paragraphs above do not give the right answer follows from what we have concluded for the actual propositional attitude or attitudes for which $<MOD,K>$ and $<MOD',K'>$ are being offered as descriptions: Even when the two constituents describe the content of the attitude they target exhaustively, their descriptions may be non-exhaustive, and moreover non-exhaustive to different extents, and by virtue of that different in content identity. Perhaps it would be unnatural to describe a's attitude at t with less detail than at t'. But even if we would be able to exclude such complex attitude attributions for reasons of unnaturalness, the problems registered in the preceding paragraphs for MSDs as mental states would still suffice to fault our current identity criteria for MSDs as descriptions of mental states as well.

Cross-indexical identity of Entity Representations

All this has been about the inadequacy of just one small set of cross-temporal identity criteria. But what other identity criteria could there be that can be applied to MSD constituents of the form $<MOD,K>$? We do not see any hope towards any such alternative, and assume that this avenue is closed. Some other way of handling cross-temporal continuity is required, which does not rely on recognizing continuity from the forms of attitude descriptions adopted in MSDRT. Such an alternative will be the subject of the remaining

Sections 4.1 and 4.2 of Chapter 4. As transition to those parts we conclude the present section with a discussion that have been ignored up to this point: the cross-temporal identity of entity representations.

From what has been said about entity representations in previous sections of this Introduction two things should have become plain: (i) that entity representations typically have temporal extension; and (ii) that entity representations typically change in the course of their extension. The most prominent changes are in their anchor sets: each time an entity representation is involved in a recognition of its referent, testimony of this enters as a new anchor into its anchor set. So, if the size and content of their anchor sets were to be included as part of the identity conditions for entity representations, then, it might seem, these identity conditions would fail for the same reasons as the contemplated identity conditions for propositional attitudes were shown to fail for them. However, for entity representations this isn't a problem in the way it is for propositional attitudes, since our treatment of entity representations assigns each of them a unique identifier in the form of its distinguished dref. So long as it can be assumed that in any MSDRT application any two entity representations are distinct iff their distinguished drefs are distinct, then distinguished drefs can serve as their unique identifiers in that application. As an illustration we repeat the MSDs presented in Ch. 2, Section 4 of the relevant parts of H's mental state before and after his processing of (2.45) in the case where H has an entity representation for Gogol that he can use in his interpretation of S's use of *Gogol*.

$$\left\{ \begin{array}{l} \left\langle [ENT, s_H]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\ ,\ \mathcal{K}_S \right\rangle \\ \\ \left\langle [ENT, g_H]\ , \begin{array}{|c|}\hline \\ \hline \text{person}(g_H) \\ \text{Named}(g_H, Gogol) \\ \hline \end{array}\ ,\ \mathcal{K}_{Gogol} \right\rangle \end{array} \right\} \tag{2.50}$$

$$(2.51)\ \left\{\begin{array}{l}
\left\langle [ENT, s_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline\end{array}\ ,\ \mathcal{K}_S \right\rangle \\[2ex]
\left\langle [ENT, g_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{p'n}(g_H) \\ \text{N'd}(g_H, \mathit{Gogol}) \\ \hline\end{array}\ ,\ \mathcal{K}_{Gogol} \cup \left\{\begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}(s_H, \mathit{Gogol}, g_H) \\ \hline\end{array}\right\} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline e\ \ y \\ \hline e \prec n \quad \text{short-story'}(y) \quad \text{by'}(y, g_H) \\ e\text{: read'}(s_H, y) \\ \hline\end{array} \right\rangle
\end{array}\right\}$$

That the ER with distinguished dref g_H in (2.51) is the same entity representation as ER with distinguished dref g_H in (2.50) is indicated by the fact that they share the dref g_H. They are the same entity representation, although in the time between them the anchor set has been augmented with the vicarious

anchor $\begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}(s_H, \mathit{Gogol}, g_H) \\ \hline\end{array}$.

In Section 4.1 we noted that so long as distinguished drefs uniquely identify their ERs, this can be exploited by using them as their labels and we introduced the notation '$\overset{\wedge}{\alpha}$' for this purpose, where α is the distinguished dref. We could use this notation also in (2.50) and (2.51), by adding the distinguished drefs of ERs also as labels. For instance, the second ER of (2.50) might be replaced by:

$$\overset{\wedge}{g_H}\text{:}\ \left\langle [ENT, g_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(g_H) \\ \text{Named}(g_H, \mathit{Gogol}) \\ \hline\end{array}\ ,\ \mathcal{K}_{Gogol} \right\rangle$$

At this point such a move seems to make little sense, since the identifying information provided by the distinguished dref g_H is there in any case. But we will return to this matter in the next section.

4.4.1 Labeling MSD Constituents

The distinguished drefs of entity representations provide them with cross-temporal identifiers. And since they do there is no need to try and find recognizable identity criteria for ERs as MSD constituents to determine which ERs in one MSD are descriptions or manifestations of the same entity representations as which ERs in some other MSD so long as the two MSDs belong to the same application. Since the temporal continuity of entity representations is essential to the various roles they play in the mental lives of agents and in the function of MSDRT's descriptions of their mental lives, that is an indispensable asset. And it is an asset that we need not only for entity representations but also for propositional attitudes. But how do we go about getting uniquely identifying labels for the latter?

Let us focus once more on the mental states that the function MS_M assigns to agents at indices in a model M. Recall that for any pair $<<w,t>,a>$ of an index $<w,t>$ and an agent a for which MS_M is defined, the value it returns is a pair $<MS_1,REF>$ consisting of an MSD MS_1 and an assignment REF. Suppose now that the constituents of MS_1 are propositional attitude and entity representation constituents that have duration, and that these will therefore also belong to the mental state of a at other indices. But at those other indices they may have different properties and stand in other relations to each other, so that their manifestations will be different there. But still, whether recognizable from those descriptions or not, there will be a fact of the matter whether a constituent $CONST$ of the mental state of a at $<w,t>$ is or is not the same as a constituent $CONST'$ of the mental state of a at $<w',t''>$. In view of the discussion in the last section, and our inability to identify this trans-indexical sameness relation in terms of features that can be defined in terms of the constituents of the MSDs MS_1, there is nothing for it but to accept trans-indexical sameness as a primitive.

This means that a model M for MSDRT has to provide even more information than we assumed so far. For each agent a in the domain of MS_M – that is, each a such that for some $<w,t>$, $MS_M(<<w,t>,a>)$ is defined – we want information for all constituents $CONST$ of mental states of a at the different indices $<w,t>$ for which $MS_M(<<w,t>,a>)$ is defined in which other mental state of a it is also a constituent. That is, each such constituent $CONST$ determines a kind of 'worm' – a partial choice function $cf(a,CONST)$ on the indexed family of those sets MS_1 such that for some index $<<w>$, $<MS_1,REF>$ belongs to $MS_M(<<w,t>,a>)$ (with REF the assignment of entities in M to distinguished drefs of entity representations in the mental states of a).

In more formal detail:

First, define for any indexed family of sets $X_{i(i \in D)}$, with i ranging over the index set D, the *partial choice functions over* $X_{i(i \in D)}$. A *partial choice function over* $X_{i(i \in D)}$ is a partial function on D that selects for each i for which it is defined an element from X_i.

Second, let DOM(MS_M,a) be the set of all indices $<w,t>$ of M such that $MS_M(<<w,t>,a>)$ is defined and let $MS_1(a)$ be the following indexed family of sets over DOM(MS_M,a):

Third, for each $<w,t> \in$ DOM(MS_M,a) let $MS_1(a)(<w,t>)$ be the MSD MS_1 such that $<MS_1,REF> = MS_M(<<w,t>,a>)$. Thus $MS_1(a)$ is an indexed family of sets, with indices from the set DOM(MS_M,a).

A partial choice function cf over $MS_1(a)$ can be thought of as connecting MSD constituents at different indices with each other as manifestations at those indices of a single constituent of a's mind, which is part of a's mind at all and only those indices that form the Domain of cf. Or, to make the same point, if $CONST$ is a constituent of a's mental state at some index $<w,t>$, then $CONST$ can be seen as the manifestation at $<w,t>$ of a certain partial choice function cf over $MS_1(a)$ for which $cf(<w,t>) = CONST$.

Our proposal is to use partial choice functions over $MS_1(a)$ to represent constituents of a's mental state as cross-indexical continuants. That is, we revise our definition of a model M for MSDRT by modifying the function MS_M from one which returns as values pairs of the form $<MSD,REF>$ to one that returns triples, of the form $<MSD,REF,CONTIN>$. Here $CONTIN$ (for 'continuants') is a function that maps each agent a on a set $CONTIN(a)$ of partial choice functions over $MS_1(a)$ with the following property:

if $<w,t>$ is an index in DOM(MS_M,a) and $CONST$ a constituent of $MS_1(a)(<w,t>)$, then there is exactly one function cf in $CONTIN(a)$ such that $cf(<w,t>) = CONST$.

We repeat this revision of the definition of the notion 'model for MSDRT' in more official format. The new definition is given in (4.23). Above it, we repeat the old definition (4.1) for easier comparison.

(4.1) (Old Definition of the *models for MSDRT*)

M is a *model for MSDRT* iff M is a pair $<M',MS>$, where

(i) M' is an IH model for DRT and

(ii) MS is a partial function from pairs $<<w,t>,a>$, where $<w,t>$ is an index from M' and a an agent from M', to pairs $<MSD,REF>$, where MSD is an MSD and REF REF is a partial function from the set DIST(MSD) of all distinguished drefs of ERs in MSD to entities from M'.

(4.23)(Revised Definition of the *models for MSDRT*)

M is a *model for MSDRT* iff M is a pair $<M',MS>$, where

(i) M' is an IH model for DRT and

(ii) MS is a partial function from pairs $<<w,t>,a>$, where $<w,t>$ is an index from M' and a an agent from M', to triples $<MSD,REF,CONTIN>$, where

(a) MSD is an MSD,

(b) REF is a partial function from the set DIST(MSD) of all distinguished drefs of ERs in MSD to entities from M' and

(c) for each agent a of M, $CONTIN$ is a set of partial choice functions over $MS_1(a)$, such that for each $<w,t> \in \mathrm{DOM}(MS_M,a)$ and each constituent $CONST$ of $MS_1(a)(<w,t>)$ there is exactly one function cf in $CONTIN(a)$ such that $cf(<w,t>) = CONST$, (where 'DOM(MS_M,a)' and '$MS_1(a)$' are as defined above).[22]

There is an obvious objection against Definition (4.23): the definition fails to impose constraints on $CONTIN$ that it should impose. The objection can take different forms, by bringing up different constraints. Here we consider a few of the most obvious ones, without any claim to completeness.

The most obvious of all is that where Definition (4.23) speaks about constituents of mental states it makes no distinction between propositional attitudes and entity representations. By not making the distinction it in principle allows for continuants that manifest themselves at some indices as propositional attitudes and at others as entity representations. But such 'continuants' are evidently absurd. Some of our mental state constituents are attitudes and some are entity representations, but no constituent is ever of the

[22] It is easily verified that the choice function set $CONTIN(a)$ defines an equivalence relation $\approx_{CONT}$ on the set $MS_1(a)$: if $CONST$ and $CONST'$ from $MS_1(a)$, then $CONST$ $\approx_{CONT}$ $CONST'$ iff there is a cf in $CONTIN(a)$ such that for some indices $<w,t>$ and $<w',t'>$, $CONST = cf(<w,t>)$ and $CONST' = cf(<w',t'>)$. But note well: this is an equivalence relation of a very special sort.

one kind at one index and of the other kind at another index. In other words, one constraint that must be imposed on $CONTIN$ is that given in (4.24.a). Furthermore, given our assumption that ERs are uniquely identified by their distinguished drefs, we also want the constraint in (4.24.b).

(4.24)a. Suppose that M, a and $CONTIN$ are as in Definition (4.23). Then for all cf in $CONTIN(a)$:

If for some $<w,t>$ in Dom(cf) $cf(<w,t>)$ is an ER, then for all $<w,t>$ in Dom(fc) $cf(<w,t>)$ is an ER.

b. If for some $<w,t>$ in Dom(cf) $cf(<w,t>)$ is an ER with distinguished dref α, then for all $<w,t>$ in Dom(fc) $cf(<w,t>)$ is an ER with distinguished dref α.

Note that (4.24) entails that when a choice function cf is such that for some $<w,t>$ $cf(<w,t>)$ is a propositional attitude, then that will also be the case for all $<w,t>$ for which cf is defined.

But for propositional attitudes further constraints come to mind as well. They are of two kinds. The first are consistency constraints on Mode. The second kind are constraints on content.

We noted earlier that for languages with sets of finely discriminating Mode Indicators it is possible that the same continuant will have manifestations with different Mode Indicators. But it would seem that this kind of variation is quite limited. A belief continuant may vary from one index to the next in strength, and changes in the Mode Indicators of its manifestations may witness that. But on the other hand an attitude that, say, is a belief at one index and a desire at another is more difficult to make sense of. We should, however, not be too hasty with such judgments. Suppose A has had a desire for some time, go and visit Venice. After some time this comes to pass. At this point A's desire has to be fulfilled. She knows that and from then on has the belief that she has been in Venice – a belief that, modulo temporal shifting, has the same content as her earlier desire. There is a sense in which in this and similar cases the desire turns into the knowledge that its content has come true. But even if there is some kind of continuity in such cases between desire and belief, it seems odd to us to claim that A's earlier persists beyond its known fulfillment, but no as the belief that what was desired has come to pass. Whatever the deeper reason for this resistance, it seems firm enough to want to rule out attitudes that involve such drastic changes in their local manifestations.

Rather than radically excluding Mode Indicator variation in the manifestations of a single attitude (by insisting on a general principle according to which whenever $cf(<w,t>) = <MOD,K>$ and $cf(<w',t'>) = <MOD',K'>$, then $MOD = MOD'$), it seems better to proceed cautiously here by separately listing pairs of Mode Indicators that are incompatible in the sense we are discussing. For instance, one might adopt the following constraints, according to which BEL is incompatible with DES and with INT:

(4.25)a. Suppose that M, a and $CONTIN$ are as in Definition (4.23). Then for all cf in $CONTIN(a)$:

If for some $<w,t>$ in Dom(cf) $cf(<w,t>) = <BEL,K>$, then for no $<w',t'>$ in Dom(cf) $cf(<w',t'>) = = <DES,K'>$ for any K'.

b. If for some $<w,t>$ in Dom(cf) $cf(<w,t>) = <BEL,K>$, then for no $<w',t'>$ in Dom(cf) $cf(<w',t'>) = <INT,K'>$ for any K'.

But what about desires and intentions. Bratman (Bratman 1987), as perhaps the first among many others, has stressed the importance of not confusing intentions with desires. But should this be understood as entailing that sometimes at least what starts out as a desire may turn seamlessly into an intention: it is still the same attitude but now with different commitments associated with it? Here the intuitions of speakers without philosophical convictions may not be quite so clear. It may still be a defensible policy to come down one way or another in a case like this one. But when we do, we must keep in mind that the decision may be about the theory of desire, intention and action, and not just about the use of cognitive categories by the linguistically educated but philosophically untutored or uncommitted.

There are also constraints on the content of propositional attitudes. When one manifestation of a propositional attitude continuant cf is of the form $<MOD,K>$ and K' is very different in content from K, then there should be no manifestations of cf of the form $<MOD,K'>$. The most extreme, and hence most indubitable, instances of this are those where K' contradicts K and especially those where the contradiction is openly displayed by representational form, as when K' is the negation of K:

$$K' = \begin{array}{|l|}\hline \\ \hline \neg\ K \\ \\ \hline\end{array}$$

But how different must K' be from K for it to right to exclude continuants some of whose manifestations have content K while others have K'? We do not know at this point whether there is much about this question that can be established for certain. This may be an area where some legislation would not go amiss: It may be useful to develop alternative versions of MSDRT which differ in what constraints on attitude continuants they adopt.

A very different kind of constraint has to do with the Domains of the functions cf. This is a matter that arises most directly in relation to time. Suppose that the continuant cf is defined for the indices $<w,t>$ and $<w,t'>$ and that $t \prec t'$, and suppose that t'' is between t and t'. Shouldn't cf also be defined for $<w,t''>$ in this case? We feel a strong inclination to accept this constraint: Once an attitude or entity representation has ceased to be part of the agent's mental state, there is no way of reviving it in the strict sense. You may have a belief and then give it up, And then, at some yet later time, you form a belief with the same content. There is a sense, of course, in a situation of this kind in which you can be described as having returned to your earlier belief. But is it the same belief, qua continuant, that you had earlier? We think not. But others may differ on this point and we leave this question as one for further debate.

We do not dare say any more about constraints on continuants until more work will have been done on this question.

Referring to Continuants by means of Labels

In the new models for MSDRT it is the functions cf in $CONTIN$ that are the 'real' constituents of mental states. And in many applications it is to them that our descriptions of mental states will have to refer. Those references, however, will still take the form of MSDs as descriptions of the mental states of agents at particular indices $<w,t>$; it will be the manifestations of propositional attitudes and entity representations at particular indices that those descriptions are made up of. It is just, to repeat, that some descriptions involve MSDs at more than one index and in those cases it is sometimes important that the description makes clear which constituents of the different MSDs are manifestations of the same continuants.

The device we will use for such cross-references between MSDs is that of labeling the constituents of the MSDs that are used to describe agents' attitudes and the mental states to which those belong. As labels we use, as a default, symbols that always contain the letter 'l'; and the labels will precede the old constituent descriptions, separated from them by a colon. For example, the MSDs for H before and after his interpretation of (2.45) might now look as in (4.26.a,b).

(4.26)a.

$$\left\{ \begin{array}{l} \left\langle l : [ENT, s_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\ ,\ \mathcal{K}_S \right\rangle \\ \\ \left\langle l' : [ENT, g_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(g_H) \\ \text{Named}(g_H, \mathit{Gogol}) \\ \hline \end{array}\ ,\ \mathcal{K}_{Gogol} \right\rangle \end{array} \right\}$$

b.

$$\left\{ \begin{array}{l} \left\langle l : [ENT, s_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array}\ ,\ \mathcal{K}_S \right\rangle \\ \\ \left\langle l' : [ENT, g_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{p'n}(g_H) \\ \text{N'd}(g_H, \mathit{Gogol}) \\ \hline \end{array}\ ,\ \mathcal{K}_{Gogol} \cup \left\{ \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: ref}(s_H, \mathit{Gogol}, g_H) \\ \hline \end{array} \right\} \right\rangle \\ \\ \left\langle l'' : BEL, \begin{array}{|c|}\hline e \quad y \\ \hline e \prec n \quad \text{short-story'}(y) \quad \text{by'}(y, g_H) \\ e\text{: read'}(s_H, y) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

The labels in (4.26.a) and (4.26.b) make explicit what was intended all along: the entity representations in (4.26.b) are the same as the identically labeled entity representations in (4.26.a). But the second one, labeled l', has changed on account of what happened between the time of (4.26.a) and (4.26.b) (H's processing of (2.45)), whereas the first has not.[23]

[23] There is one exception to this labeling convention. It concerns the labeling of entity representations. In applications of DRT and MSDRT it has been a practice to use a new discourse referent symbol each time a new discourse referent is introduced. The effect of this is that no two ERs ever have the same distinguished dref, so within the given application an ER's distinguished dref is a unique identifier of it. In footnote 5 the notational device was introduced of putting a hat over the distinguished dref of an ER to form a term to refer to the ER (as oppose to the entity represented by the ER). We can make use of these terms also as labels for the purpose described in the present section. So for instance, the first ER of (4.26.a) and (4.26.b) could also have been given as:

There ought to be a close connection between the use of labels exemplified in (4.26.a,b) and the new MSDRT models (those defined in (4.23)): the labels are ways of referring to the choice functions in the sets $CONTIN(a)$. This relation can be made explicit by modifying the correctness definition for MSDs in a way that reflects this connection. To this end we treat the labels for labeled MSD constituents as discourse referents of a special kind. As things stand, they are special for one thing in that no provision has so far been made for their binding. So we will have to make provisions for this. And for that we need to distinguish between different settings. We will assume for now that labels only occur as labels of MSD constituents. That is, they only occur in MSDs. First consider the setting in which MSDs can only occur as arguments of *Att* and thus occur in well-formed DRSs only as parts of *Att*-Conditions. In this case each MSD, and thus also each label occurrence within it, will always be in the company of the dref, say a, that fills the agent slot of the *Att*-predication. In the evaluation of the DRS K in which this *Att*-predication occurs in a given model M, a will be assigned an entity from M at some point; let f be the assignment that assigns this entity to a. We stipulate that when this assignment takes place, it is to be accompanied by an assignment of choice functions from $CONTIN_M(f(a))$ to all the labels occurring in MSDs that are part of *Att*-Conditions in K in which a occurs as agent argument. (This extension of f to the labels in these MSDs can be seen as an interpretation of how these labels are used to refer to continuant constituents of the agent $f(a)$.)

But how and when can the agent arguments of *Att*-predications be assigned in the course of DRS evaluation? Where can the assignment of an agent from M to such a dref a occur? For the DRSs K considered here there are two possibilities: (i) a occurs in the Universe of K or of some sub-DRS of K. In this case making an assignment to a is part of the standard evaluation procedure for DRT. (ii) a is the distinguished dref of some ER occurring as part of some MSD. But then this MSD will occur as part of an *Att*-predication and thus be interpreted as the mental state of the agent of M that will get assigned to the agent argument of this *Att*-predication. Suppose that this second agent argument is a' and that it has been assigned the value $f(a')$. Then the value of a will be determined by the function REF from the value $<MS_1,REF>$ assigned to $f(a')$ by MS_M. Ideally these observations should be replaced by a fully explicit statement of the verification conditions of DRSs with labeled

$\left\langle \hat{s}_H\colon [ENT, s_H]\ ,\ \begin{array}{|c|}\hline \\ \hline \text{person}(s_H) \\ \hline \end{array},\ \mathcal{K}_S \right\rangle$. Of course, when ERs are labeled in this way, then the dref in the label must always be the same as the one following ENT.

MSD constituents. But that would be a lot of further work, which at this point doesn't seem worth the effort.

The second general setting is the communication-theoretic approach to semantics on the basis of MSDRT. That is the setting of the one example that we have so far given of MSDs with labeled constituents, in (4.26). The problem here is that we haven't yet extended our model-theoretic semantics for MSDRT to this setting at all so far. Actually, such an extension is fairly straightforward. For instance, consider the analysis of the *Gogol* sentence (2.45) in Section 2.4. We can subsume this analysis under the model theory of the present section by treating S and H as agents in a model M in which S's utterance is an event occurring in some world w within some temporal interval t and in which these agents are in their respective mental states in w during this period. The mental state descriptions attributed to S and H in Section 2.4 can then in principle be evaluated for correctness in M, using the correctness definition developed in Section 2. (Though that possibility is of course a merely abstract one, so long as we haven't been given more concrete information about what M has to say about the mental states of S and H.) As far as Section 2.3 is concerned, all that is needed is a model M for MSDRT according to the old definition (4.1). When we now switch to attributions like those in (4.26), it will be necessary to switch to models according to the new definition in (4.23). But that transition will involve the observations described in the last two paragraphs.

A logical matter connected with labeling is to what extent adding labels changes the logic of the DRS languages thus extended. Before we say more to this matter, let us first remark on a general omission of this MSDRT introduction: Nothing has been said about the logical properties of the DRS languages of MSDRT, among them classical metamathematical properties like axiomatizability. Investigating such properties was long part of the research agenda of DRT.[24] Unlike the situation sketched in the footnote above,

[24] It is also reflected in the title of (Kamp & Reyle 1993). The book was planned as a two volume project from which in the end only the first half was published. The second volume was going to be devoted to the logic of DRS languages, along lines motivated by the intuition that it is part of natural language discourse that certain inferences from the contributions to the discourse are licensed by the forms of those contributions alone, and of which logic as it has come to us through the ages from the time of Aristotle and culminated in the definition and investigation of first order and higher order predicate logic, is the ultimate abstraction. The second volume never made it to a stage that could have been published. One part of it – an axiomatisation of first order DRS languages by axioms and rules tuned to the form of DRSs and DRS Conditions, was published as (Kamp & Reyle 1996). But this part of the project lost impetus after the authors found that when

the logical investigation of fragments of the DRS languages of MSDRT that are described and used in this Introduction may hold more promise. Among these fragments there is one that subsumes the large majority of the DRSs presented in this and the preceding sections: DRSs in which *Att*-predications occur only as Conditions in the Condition Set of the main DRS (and thus not in the scope of negation or other logical operators).

In fact, some more thought may be needed about how to deal with logically embedded *Att*-predications. Our proposal about the verification conditions of *Att*-Conditions has been guided by what it is for them to be verified, and not by what conclusions are to be drawn for verification definition for the containing DRS when the verification conditions for an *Att*-Condition are not satisfied. And with regard to this question we envisage significant differences between *Att*-Conditions with MSDs with and without labeling. Adding labels tends to create new instances of logical consequence, which disappear when the labels are removed from premises and conclusion. What new instances of the consequence relation are created by labeling will depend, however, on the constraints on the $CONTIN$ constituents of the models.

In this document we have nothing more to say about the logical properties of DRS languages for MSDRT. All this is among the many topics for possible later work that we have been mentioning as we have been going along.

4.4.2 Labels as fillers of argument slots

In the last section labels were introduced as devices for the cross-indexical identification of mental state constituents. There is also another use of labels in MSDRT, which was introduced in unpublished work much earlier. It has to do what in MSDRT is known as *self-reflection*.

We consider self-reflection in MSDRT an important topic in its own right, which deserves at a minimum its own separate section. That we devote only a sub-sub-section to it in the present document is a way of saying that the section that self-reflection deserves is missing here, and we are bringing it up here as an alternative reason for the use of labels.

DRT is extended to plurals one moves into the domain of second order logic, and they were unable to find interesting fragments of plural DRT with tractable consequence properties (axiomatizability or decidability). Some of the more conceptual questions connected with the project are still without an optimal formulation today.

The term 'self-reflection' is used in MSDRT to refer to the thought processes of an agent that are about her own thoughts. We do this a lot of the time. We wonder how we ever got seduced into a belief that we now think is offensive and absurd – when one of us was a child growing up in the Netherlands he believed it was an injustice that 'his country' had 'lost' the 'Dutch Indies'; the author has repeatedly asked himself later how he could have been so daft. We think it is ridiculous for us to be afraid of spiders, but we are afraid of them no less. We often wish that we didn't have a desire that we do have, like the former smoker, who is craving for a cigarette and wishes he wasn't. But these are only the simplest kinds of examples of their kind.

Self-reflection has much in common, obviously, with 'second order' attributions, in which one agent A attributes to another agent B an attribution that B makes to some further agent. For such second order attributions we have a mode of representation: A's attribution to B takes the form of an *Att*-Condition whose agent argument refers to B and whose MDS argument describes the attribution by B that A attributes to him. And often, though not necessarily, the attribution that A makes to B also involves an Entity Representation that A has for B.[25] Part of this representational machinery is intended to capture the external aspects of such attributions: the attribution is to an agent external to the attributor and so is what is being attributed to him – whether that is a thought of that agent or some other property. Self-reflection is different from this in that our access to our own thoughts is direct in a way that our access to the thoughts of others never is. Our access to our own thoughts is direct in a way that is reminiscent of our access to our own selves. To the authors of the present document this suggests that agents should have devices for referring to their thoughts that are special in the manner of the self-reflective discourse referent i. Labels are our implementation of this idea: in MSDRT's representations of self-attributions the self-attributing agent is represented as referring to her own thoughts by means of labels; the labels are to be seen as signaling her direct access to those thoughts, so that she can make them into arguments of the predications that are involved in her second order thoughts about them.

Implied by this last sentence is that the use of labels in self-reflections extends syntactically beyond the use of labels in the last section: labels do not longer occur just as labels but also as arguments of predicates. The following

[25] An example where this is not so is when A attributes a certain attitude to a group of people, in the sense that everyone in that group has the attitude, and where B belongs to that group but A doesn't know that, and may not even know B at all.

examples gve an impression of how this works.

The first example is an MSD for the thought of some ancient Babylonian, who asks himself whether the Morning Star might not be the same as the Evening Star. We refer to the Babylonian quasi-anonymously as 'A' and assume, no doubt falsely, that 'Mosta' and 'Esta' are the names by which he referred to the Morning Star and the Evening Star. The attitudinal mode of A's thought, that of querying, or wondering, is represented in MSDRT by the mode indicator 'WON'. The MSD in (4.27) gives the relevant part of A's mental state, consisting of ERs for the Morning Star and the Evening Star and his query about them. The symbols used for the labels for the two ERs in (4.27) have been chosen in accordance with the convention described in footnote 23 of the last section.

$$
(4.27)\quad \left\{ \begin{array}{l}
\hat{m}\text{:} \left\langle [ENT, m], \boxed{\begin{array}{c} \\ \text{Named}(m, \mathit{Mosta}) \end{array}}, \mathcal{K}_m \right\rangle \\[2em]
\hat{e}\text{:} \left\langle [ENT, e], \boxed{\begin{array}{c} \\ \text{Named}(e, \mathit{Esta}) \end{array}}, \mathcal{K}_e \right\rangle \\[2em]
l : \left\langle WON, \begin{array}{|c|} \hline \\ \hline m = e \\ \hline \end{array} \right\rangle
\end{array} \right\}
$$

The next example of an attitude with a self-reflective content specification is that of an agent a who believes of one of her entity representations that she already had it the year before. The MSD for the relevant part of the mental state of a at the time when she holds this belief consists of two constituents, the ER and the belief that this ER was already part of her mental state the previous year.

$$
(4.28)\left\{\begin{array}{l}
\hat{x}\text{: } \langle [ENT,x],K,\mathcal{K}\rangle \\[1em]
l : \left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline \text{cal.year'}(t) \quad s\subseteq t \\ \begin{array}{|c|}\hline t' \\ \hline \text{cal.year'}(t') \quad n\subseteq t' \quad t \supset\subset t' \\ s\text{: EXIST}(\hat{x}) \\ \hline\end{array} \\ \hline \end{array}\right\rangle
\end{array}\right\}
$$

Here the predication ‘s: EXIST($\hat{x}$)’ is second order in the current sense of that term insofar as the second argument term, ‘$\hat{x}$’, refers to a first order MSD constituent. (The ER in the top line of (4.28).) Note that ‘EXIST’ is a genuine predicate, expressing a relation between temporal continuants and states which holds when the continuant exists throughout the state. But it can enter into higher order predications when the continuants it takes are MSD constituents.

The last example is a formal representation of the case mentioned in the introductory part of the present section: the former smoker who feels a desire for a cigarette and at the same time the ‘meta-desire’ not to have that desire.

$$
(4.29)\left\{\begin{array}{l}
l : \left\langle DES, \begin{array}{|c|}\hline s \quad y \\ \hline n\subseteq s \quad \text{cigarette'}(y) \\ s\text{: PROG(the } \hat{e}. \begin{array}{|c|}\hline \\ \hline e\text{: smoke'}(i,y) \\ \hline\end{array}\ ) \\ \hline \end{array}\right\rangle \\[2em]
l' : \left\langle DES, \begin{array}{|c|}\hline s' \\ \hline n\subseteq s' \\ \neg \begin{array}{|c|}\hline s'' \\ \hline s''\subseteq s' \\ s''\text{: have'}(i,\hat{l}) \\ \hline\end{array} \\ \hline \end{array}\right\rangle
\end{array}\right\}
$$

When the use of labels in representations of self-reflections was first conceived, the question whether or not the constituents of mental states are continuants wasn't an issue considered. But it wasn't denied and of course it is plain that what goes on in self-reflection much of the time would make no sense unless the objects of self-reflection, viz. the thoughts that the self-reflections are about, didn't have a certain temporal persistence. So there is no incompatibility between labels as identifiers of the objects of self-reflection and labels as devices of cross-identical identification. From now on we think of labels in MSDRT as performing both these functions.[26]

This ends our discussion of labels and with that the discussion of MSD constituents as cross-indexical and cross-temporal continuants.

4.5 MSDs and Mental States in Models

In Section 4 we assumed that mental states are MSDs. As noted, that assumption is far from self-evident. It may be true that agents have linguistically grounded representations for some of their attitudes and the assumption that those attitudes take the form of MSD constituents may be plausible too. But to claim that mental states *always* consist of such constituents is a claim of a different order, and one that doesn't have much to be said for it. Our mental state ontology shouldn't exclude a priori the possibility that the constituents of the mental states of human agents are not given as MSDs, or at least not all of them, and certainly not always.

One way to think about the ontology of mental states is by focusing on the function of their constituents rather than on their representations. As a matter of fact, the treatment of mental states adopted in Section 4 provides a useful basis for such an approach. That treatment identifies mental states as MSDs, but the correctness definition for MSDs that is its center piece relies on the *denotations* of those representations – it is the entailment and equivalence relations between those denotations that determine whether a given MSD is or is not a correct description of the mental state it targets.

[26]With an eye on the application of MSDRT that led to thie present MSDRT introduction we note temporal continuity is not only an essential feature of individual intentions, plans and their executions, the topic of (Bratman 1987), but even more so of joint intentions, planning and actions, the topics of (Bratman 2014) and the second part of (Ludwig 2016). An important aspect of many cases of shared intention are the negotiations between agents that lead to the intentions they share and the often joined planning that follows. As a rule such negotiations involve many different attitudes in the process. A correct formal description of those processes would be in even greater jeopardy without proper formal ways of handling continuant attitudes.

This observation supports the idea that a conservative conception of the ontology of mental states should be limited to those denotations: the denotations are needed for the model theory of Section 3 in any case, and no more than them is needed to implement the model-theoretic agenda of Chapter 4. So why not assume that the mental states of agents in models consist just of those denotations, and leave it as a separate topic of investigation how these denotations might be represented by the agent, in some form of mentalese (and perhaps in the form of an MSD) or by a phrase from the agent's mother tongue?

4.5.1 Jointly Localized Function Systems

It may be thought that retreating to such a more abstract ontology is the right way to go. But if we want to go this way, there is one important lesson from the earlier parts of Section 7 that should be firmly kept in mind. The denotations of MDS constituents are not isolated semantic entities. They are connected with the denotations of other constituents of the given MSD, just as the MSD constituents whose denotations they are may be connected to other constituents by referential dependence: We must be able to recapture referential dependence relations at the more abstract level of the denotations.

The denotations of content specifications of MSDs, we have seen, are functions of some kind. Furthermore, the denotations of the content specifications of an MSD in a model M are entities that can be constructed from elements of M by purely set-theoretic operations; in that sense these denotations are part of M, in that extended sense, in which it includes besides its basic constituents also all entities that can be built from those with the tools of set theory. But to repeat, it would be wrong to think of the 'denotation in M' of an MSD as simply the set of the denotations in M of its different constituents. What we need instead are what we call *systems of jointly localized functions.*

To explain at the hand of an example what we mean by this, let us go back once more to our very first example of an MSD, first presented as (2.3) in Section 1. We display it here once more.

$$
\left\{ \begin{array}{l}
\left\langle BEL, \begin{array}{|c|}\hline x \quad s_1 \quad s_2 \\ \hline n \subseteq s_1 \quad n \subseteq s_2 \\ s_1\text{: gold-coin'(x)} \\ s_2\text{: lie-in-front-of'}(x,i) \\ \hline \end{array} \right\rangle \\[2em]
\left\langle DES, \begin{array}{|c|}\hline s_3 \\ \hline n \subseteq s_3 \\ s_3\text{: have'}(i,x) \\ \hline \end{array} \right\rangle \\[2em]
\left\langle INT, \begin{array}{|c|}\hline t \quad e \\ \hline n \prec t \quad e \subseteq t \\ e\text{: pick-up'}(i,x) \\ \hline \end{array} \right\rangle
\end{array} \right\} \tag{2.3}
$$

There are three content specifying DRSs in this MSD. In the following discussion we refer to the first two of these as K_{BEL} and K_{DES}.[27] According to the terminology introduced earlier in Section 7, the first of these denotes an information state and the second and third denote *ccp*s. The information state SAT denoted by K_{BEL} in a model M can be seen as determined by K_{BEL}'s Condition Set: a function from the indices of M to sets of assignments, all of which have the Universe $\{x,s_1,s_2\}$ of K_{BEL} as their Domain.[28] The contributions made by the Conditions in the Condition Set of K_{BEL} to the information state $SAT_{M,Bel}$ K_{BEL} denotes in M are reflected in the assignment sets $SAT_{M,Bel}(<w,t>)$ for the different indices $<w,t>$: which assignments are in any of these sets and which are not. This is the 'proposi-

[27] We won't have anything to say in what follows below about the content specification K_{INT} of the intention of (2.3). What could be said about K_{INT} would be a redundant duplication of what will be said about K_{DES}.

[28] There is a sense, of course, in which the information state determined by K_{BEL} also depends on the contributions made by the special drefs n and i. But these contributions are fixed by the special regime that governs the denotations of these drefs. Recall in this connection that evaluations of DRSs that are parts of Logical Forms of sentences or discourses are always evaluations of the Logical Forms of *utterances* of those sentences and discourses. In particular, the MSD (2.3) will in practice be evaluated as part of a Logical Form at indices $<w,t>$ of M at which the represented sentence or discourse was uttered. That fixes the entities represented by n and i as the time t of that utterance and its speaker. These remain fixed in the evaluations of all the constituent parts of the Logical Form, including the content specifications of MSDs occurring in *Att*-predications this Logical Form may contain. **N.B. Probably this observation should come earlier. But where?**.

tional' dimension of information states, which is partly reflected also in the propositions that can be derived from them, and which distinguish only between empty and no-empty sets $SAT_{M,Bel}(<w,t>)$.

The entity determined by the Condition Set of K_{DES} doesn't determine an information state in M but a context change potential, a function from information states to information states (and one that is not reducible to an information state in the sense explained towards the end of Ch. 4, Section 2.1). Context change potentials are more complex entities than information states in that they involve two 'parameters' instead of one: (a) the set of those information states for which they are defined, and (b) what information results when the *ccp* is applied to such an information state. (Information states only involve the second of these two parameters; thought of as *ccp*s, they are the ones that are defined for all information states as inputs.) Of these two parameters, (a) is determined by those drefs in the Condition Set of K_{DES} which are of type (iv) in the classification (4.2). In the present case this is just the dref x. So, in the terms first introduced in Section 2.1 of this Chapter, $PRES$ is the singleton set $\{x\}$ in this case, while the set D is the Universe $\{s_3\}$ of K_{DES}. It follows that the information states SAT for which the *ccp* denoted by K_{DES} in M is defined are those for which ArgD(SAT) $\supseteq \{x\}$ and ArgD(SAT) $\cap \{s_3\} = \emptyset$. Furthermore, the propositional dimension of the *ccp* CCP_{DES} denoted by K_{DES} is manifest through the differences between the propositions derived from the information states $CCP_{DES}(SAT)$ and the propositions derived from the input state SAT, for the various information states to which CCP_{DES} can be applied. Among the infomation states for which CCP_{DES} can be applied there is in particular the one denoted by K_{BEL}; recall the discussion of the primary information states associated with *ccp*s in Section 2.2.

This much should suffice as a reminder of how the denotations of different propositional attitude constituents of MSDs can hang together. But the reminder is also meant to bring out an aspect of the denotations of the content specifications of MSD constituents that we did not comment on in our earlier discussions of this or other example: There is a sense in which all such denotations involve propositional functions, that these propositional functions are often connected with each other through the mechanisms of referential dependence, but that for them the connections take a form that has not yet been articulated. We can see more clearly what these connections are like, when we focus on the Condition Sets of the content specifying DRSs of MSDs, along the lines of the discussion above.

One way to explicate these connections is in the abstract and general terms of systems of jointly localized functions. We assume that every function is in part characterized by its Argument Domain ArgD(f). ArgD(f) corresponds roughly to the argument sets that are specified by lambda terms in Standard Type Theory. In Type Theory functions are distinguished in terms of their logical types. A function f of type $<\alpha_1<\alpha_2< \ldots <\alpha_n.\beta>..>>$ is one that returns a value of type β when applied to arguments of the types $\alpha_2< \ldots <\alpha_n$ in that order. ArgD(f) then corresponds to the set $\{\alpha_1,\ldots, \alpha_n\}$. The correspondence is not perfect in that (i) ArgD(f) doesn't come with a fixed order and (ii) ArgD(f) doesn't impose any restrictions of the types of the arguments that can fill its different positions. As a matter of fact, the functions that matter here come with clear constraints on the sorts of arguments that can fill particular argument positions of certain functions. For instance, some argument positions will be reserved for the indices $<w,t>$ of models. But our handling of the sortal restrictions on argument positions in what follows will be ad hoc; we will state the restrictions as we go along and it seems helpful to do so.[29] Applying a function f takes the form of combining it with an assignment of admissible arguments to ArgD(f).

A function f can not only be applied to real arguments, it can also be *instantiated to argument terms.* Such instantiations take the form of assignments of argument terms (expressions from some language L) to the elements of ArgD(f). We call such assignments *localizations (in L).* When there is a designator **f** for f in the language L of a given localization *loc* of f, then we can form the *function term* of L by combining **f** and *loc* into the expression '**f**(*loc*)'. In what follows we are interested in localizations of a special sort, so-called *canonical localizations.* The instantiating argument terms of a canonical localization are always discourse referents taken from an infinite special purpose set of drefs, set aside for this purpose. We refer to this set as $CANDREF$. From now on here we only consider canonical localizations.

A localization of a function f determines an application of f when the terms τ_i of the localization designate entities $[[\tau_i]]$. The assignment involved in the application will then be the one which assigns to each argument position in ArgD(f) not its localization value τ_i but instead the designation $[[\tau_i]]$ of τ_i. In particular, a canonical localization *loc* of f can be turned into an application by an assignment a of entities to the drefs occurring in *loc*. In this case the

[29]In other contexts it will be useful to assume that the members of ArgD(f) come with their sortal restrictions and then it will be more useful to specify explicitly and from the start what the sortal restrictions are. The restrictions may then be just to types in the sense of Standard Type Theory, but they may also involve other kinds of sortal constraints.

assignment that determines the application of f is the composition $a \circ loc$ of loc and a.

One of the possible effects of localizing a function f is that some of the argument positions in ArgD(f) are bound together. Localizations need not be 1-to-1. And when they are not, they will insert the same term into more than one position. This means that every application determined by a given localization loc of f will insert the same argument into the positions that share the same term when f is localized by loc. (Argument positions localized to different terms may also get the same argument in an application determined by the localization, but they need not be.)

What matters for our reconstruction of the information conveyed by an MSD are *systems of functions* with a *joint localization*. The *joint localization* of a set $\mathcal{F}$ of functions is a simultaneous instantiation of all the functions in $\mathcal{F}$ with terms from some given language L; a *canonical joint localization* is a joint localization in which all terms are drefs from $CANDREF$. Such a joint localization is not only able to impose coreference on argument positions of the same function but also on argument positions of different functions in $\mathcal{F}$. It is the coreference links between argument positions of different functions that are crucial to the reconstruction of MSDs as structures of referentially dependent contents. From here on our focus will be on systems of functions with canonical joint localizations.

However, just by itself joint localization won't give us all we need. Joint localization is a symmetric relation between functions, but referential dependence is an asymmetric relation between MSD constituents. To turn joint localization into a relation that can capture referential dependence, and which must therefore also be asymmetric, we need one further distinction that can be applied to systems of jointly localized functions. This is a distinction that corresponds to the distinction introduced in Section 2.1 between the sets $PRES$ and D associated with the Condition Sets of content specifying DRSs. Our reconstruction of referential dependence assumes that this division comes as part of the canonical joint localizations of function systems $\mathcal{F}$: For each f in $\mathcal{F}$ the canonical joint localization of $\mathcal{F}$ comes with a division of the localization of ArgD(F) into two parts, its $PRES$ part and its D part.[30]

[30]When discussing the free drefs of the Condition Sets of content specifying DRSs in Section 2.2 we noted that besides the drefs in the $PRES$ part and those in the D part there are also those of type (ii) in the sense of the classification in (4.2). These drefs get their values higher up in the evaluation of the MSDs to which these DRSs belong. In the systems of jointly localized functions that we propose here as part of a possible ontology

This division, moreover, can be extended to the argument positions of f: By stipulation an argument position in ArgD(f) belongs to the *PRES* part of ArgD(f) iff it is filled with a dref belonging to the *PRES* part of the *PRES-D* division of localization of f, and likewise for the D-part of ArgD(f).

We assume that the *PRES-D* divisions of the drefs involved in the localizations of the individual functions in a system $\mathcal{F}$ is a third component of such a system, a function *PRS-D* that maps each f to the pair of disjoint sets <*PRS*,*D*> that together make up the set of drefs that fill argument positions of f. So formally, the systems that we propose as replacements for MSDs as mental states of agents are triples <$\mathcal{F}$,*loc*,*PRS-D*>. But not all such triples are systems that can serve as mental states. Here are a number of constraints they should satisfy, all of them inspired by earlier observations about referential dependence.

(a) The *PRES* part of any function f consists of instantiated argument positions that are linked by coreference with an (identically instantiated) argument position of at least one other function g in $\mathcal{F}$ such that this argument position belongs to the D part of g. (The instantiated argument positions in the D part may be linked by coreference to argument positions of other functions too, but they need not be.)

(b) When an argument position from the D part of f is linked by coreference to an argument position of some other function g in $\mathcal{F}$, then that argument position of g must belong to the *PRES* part of the instantiated argument positions of g.

It is not difficult to verify that when <$\mathcal{F}$,*loc*,*PRES-D*> satisfies the conditions (a) and (b), then the following relation RefDep is asymmetric, where RefDep is defined thus:

for f and $g \in \mathcal{F}$, f RefDep g iff (i) an argument position of f is linked by coreference to an argument position of g and (ii) this argument position belongs to the Dom part of ArgD(f).

In addition we stipulate that the relation RefDep is well-founded on $\mathcal{F}$ and we also restrict our attention to cases where $\mathcal{F}$ is finite. These assumptions guarantee that RefDep has the structural properties imposed on referential dependence between the propositional attitude constituents of MSDs (see

for mental states such drefs play no part. For the functions f of these systems the *PRES* part and the D part make up all the argument terms involved in the localization of f.

Section 2.1 of this Chapter).

So far, all we have said about the functions in systems $<\mathcal{F}$,*loc*,$PRES$-$D>$ had to do with the argument positions of the functions in $\mathcal{F}$, the arguments to which they can be applied and argument terms that can fill them. Equally important is what kinds of values the functions return when they are applied to suitable arguments. Here too, the stipulations that follow are inspired by what we have been saying about mental states as MSDs, and more specifically about the denotations in models of the content specifications of those MSDs, the information states denoted by content specifications with empty $PRES$ part and the *ccp*s denoted by content specifications with non-empty $PRES$ part. A similar difference will be central also to the functions of the systems $<\mathcal{F}$,*loc*,PRS-$D>$ that we want to propose as a possible non-representational ontology for mental states. But in relation to such a system $<\mathcal{F}$,*loc*,$PRES$-$D>$ we need to distinguish between the functions f in the set $\mathcal{F}$ and the functions that are imposed by the division PRS-D on the localizations of those functions by *loc*. The former are all propositional functions of sorts – functions defined on the set of indices of a given model which return for each index a function that maps assignments of entities to their ArgD sets to the set $\{T, \perp\}$ of the two truth values T (True) and $\perp$ (False).[31] But these propositional functions can be turned into other functions by the divisions on their argument sets imposed by $PRES$-D. When the $PRES$ part of a function f from $\mathcal{F}$ is empty, then there is no change. The 'localized function' determined by f is just f. But when the $PRES$ part of f is non-empty, then there is a real change. In this case f is turned into a partial function f' from propositional functions to propositional functions. f' is defined on those propositional functions g for which (i) their D part includes the $PRES$ part of f and (ii) D part is disjoint from the D part of f. And the result h of applying f' to such a function g is in essence the conjunction of the propositional information in g and in f. More precisely, for any index $<w,t>$, $h(<w,t>)$ is a set of assignments that is defined on the set of drefs filling argument positions of either g or f (or both) and which for each such assignment a returns the truth value $((g(<w,t>))(a_g)$ & $((f(<w,t>))(a_f)$, where a_g and a_f are the restrictions of a to the sets of argument terms of g

[31] Such functions are simple recodings of information states as defined in Section 2. Information states map indices to sets of assignment functions, intuitively the sets consisting of those assignment functions that 'satisfy' the information state. The propositional functions spoken of here map indices onto the characteristic functions of those sets. They are if you like, just our earlier information states, specified by a slightly different but trivially equivalent definition.

and f, respectively.[32]

This brings us to the first part of the central proposal of the present section: The propositional attitude part of the mental state of an agent in a model consists of pairs $<MOD,pf>$ where the pfs are propositional functions or functions from propositional functions to propositional functions determined by a system $<\mathcal{F},loc,PRES\text{-}D>$ of the kind described above.

But mental states do not consist just of propositional attitudes, they also consist of entity representations. There has to be a second part to the proposal which deals with those. And here too our proposal should conform to the general aim of this section to provide a non-representational alternative to mental states as MSDs.

In the light of all that has been said about entity representations as MSD constituents this over-all purpose seems to leave us with a number of options. A minimal option is that according to which an entity representation should be a device which does no more than (a) capture which entity is represented by the entity representation and (b) contribute the entity it represents to predications elsewhere in the mental state (by inserting its distinguished dref into argument positions of propositional functions). One identification of entity representations that fits this bill is that entity representations are pairs $<[ENT,\alpha],\mathbf{d}>$, where α, an element of $CANDREF$, is the distinguished dref of the entity representation and **d** is the entity represented. This minimal proposal is able to capture the role of entity representations in the singularity of thoughts, but it doesn't account for much else. No descriptive information about the represented entity, no room for entity representations that fail to represent, no reinforcement of the representation relation through recognition or the derailments caused by false recognitions. This minimal proposal could be extended with features, that correspond to further aspects of ERs with the form assumed throughout the present document. But in this first pass we will stick to the proposal we have made, leaving the explorations of less arid alternatives for us or others to explore at some later time.[33]

---

[32] Note that the function f' should not only be defined for the localized functions g in $\mathcal{F}$ that satisfy the constraints – that the $PRES$ part of g is empty, that the $PRES$ part of f is included in the D part of g and that the D parts of f and g are disjoint – but also for other propositional functions satisfying these constraints, which do not belong to $\mathcal{F}$, but also involve canonical localization. Among these are among others propositional functions that result from applications within $\mathcal{F}$, but which do not belong to $\mathcal{F}$ as given.

[33] An alternative way of adding the effects of entity representations is to define a subset $DISTDR$ (for 'Distinguished Discourse Referents') of the drefs occurring in loc as the set

Adopting the proposal of the last paragraph for the identity of entity representations, we come to the following alternative proposal for the mental states of agents in models:[34]

(4.30)The *mental state of* an agent a in a model M at an index $<w,t>$ is a pair consisting of:

(i) an infinite set $CANDREF$ of discourse referents and a system $<\mathcal{F},loc,PRS\text{-}D>$ of jointly localized propositional functions with PRS-D divisions; and

(ii) a set of constituents which are either

(a) of the form $<MOD,pf>$ where MOD is a Mode Indicator and pf is the propositional function or function from propositional functions to propositional functions induced by PRS-D-induced from some member f of $\mathcal{F}$ or

(b) of the form $<[ENT,\alpha],\mathbf{d}>$, with $\alpha \in CANDREF$ and $\mathbf{d}$ an element from M.

This completes our proposal for a non-representational ontology for mental states. Given the care we have taken to reconstruct the structure of mental states that is presupposed by a formal definition of correctness of MSDs as descriptions of mental states, it should be possible to recast the definition of Section 3.1 – and with it the verification definition of Section 3.2– for the mental states of (4.30). This will be the task of the next section.

---

consisting intuitively speaking of the distinguished drefs of entity representations that are constituents of the given mental state and to assume an assignment $ASS[DISTDR]$ of the drefs in $DISTDR$ to entities from the model M. That is, the mental states of agents in models are on this alternative conception mo longer given by just the combination of a structure $<\mathcal{F},loc,PRES\text{-}D>$ and a set of pairs $<MOD,pf>$, but will have as an additional component a function $ASS[DISTDR]$ (which determines $DISTDR$ as its domain). Assignments used in applications of the functions in $<\mathcal{F},loc,PRES\text{-}D>$ to entities from M will now have to 'respect' $ASS[DISTDR]$ in that they agree with $ASS[DISTDR]$ on all drefs that are arguments to both. This guarantees that the content functions pf from the paris $<MOD,pf>$ represent propositional functions in M that are singular with respect to the values under $ASS[DISTDR]$ of the argument positions identified by drefs in $DISTDR$.

[34]This proposal ignores the indexical continuity of propositional attitudes, the topic of the preceding Section 4. To deal with continuity we can augment the present proposal with a labeling system of the sort described in that section.All that we are saying about the alternative proposal in this section is compatible with the addition of such a labeling system.

4.5.2 Defining correctness for MSDs as descriptions of the mental states of Section 4.5.1

Adapting the correctness definition of Section 3.1 to the models of the last section is not a difficult task. In fact, hardly any modifications are needed at all. As before, MSD correctness requires maps TRANS and TARGET, where now TRANS is always an injection of the set of drefs occurring in the given MSD MSD into the set $CANDREF$. For a propositional attitude constituent $<MOD,K>$ of the MSD it is now necessary that the denotation of K in M stands in the right relation (of entailment or equivalence) to the second member of TARGET($<MOD,K>$). If the denotation of K, $[[K]]_M$, is an information state, then the second component of TARGET($<MOD,K>$) should be a propositional function with an empty PRS part. Since such propositional functions are information states in disguise (and a quite transparent one at that), entailment and equivalence are the same relations of entailment and equivalence in this case that we defined as relations between information states to begin with. Likewise for information state equivalence.

If $[[K]]_M$ is a *ccp* that is not an information state, then the second component of TARGET($<MOD,K>$) – let us refer to this component as TARGET$[K]$ for easier reference – should be a function induced from a function f in $\mathcal{F}$ by a non-empty $PRES$ part. So the $PRES$ part from $PRES$-$D(f)$ should be non-empty. But this is guaranteed by the obvious adaptation of the constraint on TRANS and TARGET that is part of the correctness definition in Section 3.1, viz. that the TRANS($PRES_K$) where $PRES_K$ is the $PRES$ set determined by K within MSD, is included in the $PRES$ part of TARGET$[K]$. The further constraint on the correctness of MSD that is imposed by the *ccp* CCP denoted by K depends as before on whether MOD is EXREQ or NON-EXREQ. In the latter case the constraint is now that when IS is an information state in the Domain of CCP, PF is a propositional function in the Domain of the function from propositional functions to propositional functions TARGET$[K]$, and the information state IS' determined by PF via our trivial recoding entails IS, then this entailment is preserved by applying CCP to IS and TARGET$[K]$, to PF: the information state $CCP(IS)$ is entailed by the information state determined (in the sense of recoding) by TARGET$[K](PF)$. When MOD is EXREQ, then we need equivalence between CCP and TARGET$[K]$, but again it is straightforward to adapt the clause of the correctness definition of Section 3.1 so that it imposes the intuitively right constraint in this case too.

The constraints on correctness contributed by an ER $<[ENT,\alpha], K, \mathcal{K}>$

from MSD consists of two clauses. The first is that if TARGET($<[ENT,\alpha]$, K, $\mathcal{K}>$) is the entity representation $<[ENT,\beta],\mathbf{d}>$ of the agent's mental state, then it must be the case that TRANS(α) = β. As before, the second constraint arises in those contexts where MSD occurs as third argument of *Att*, and has to with the corresponding Links argument. If Links contains the pair $<\alpha,\alpha'>$, then assignments used to evaluate the *Att* predication and which must therefore be defined on α', must assign $\mathbf{d}$ to α'.

Nothing in this adaptation of our earlier correctness definition to this section's notion of mental states is either surprising or greatly illuminating. That it should be possible to formulate a correctness definition for MSDs and a verification definition for *Att*-predications in relation to the mental states proposed in this section that match the definitions of Sections 3.1 and 3.2 in form and over-all plausibility may have been clear enough from the details of Section 5.1, and perhaps the details of the last three paragraphs may have been a gilding of the lily that did it little good (and certainly haven't made it look prettier). Good reason to leave these details, at least for now. But whether the details of the last few paragraphs should remain or be thrown out, more formal details at this point could only make things worse.[35]

But one semi-formal comment of a general nature is still needed. One obstacle we encountered with the definitions of correctness and verification conditions in Section 3.1. and 3.2 was that in some cases we didn't have the denotations of content specifications of mental states of agents in models at the point when we needed them. We finessed the problem as best we could by using hypothetical denotations and hoping that everything would come out right in the end. But such a strategy provokes questions that we noted but that we had no serious answers to. With the mental state ontology proposed in the present Section 5 the difficulty doesn't arise. The propositional functions and functions from propositional functions to propositional functions that identify the contents of the propositional attitudes in these mental states are what they are; there is no problem here of contents that are given

[35]We may be accused of a certain incoherence in having returned in this section to the state of play reached at the end of Section 3, rather than focus on the syntax and semantics of the labeled MSDRT laid out in Section 4. Our reason has been that we haven't so far presented an explicitly worked out semantics for MSDs and *Att*-predications of labeled MSDRT, which we could have modified to fit the mental states of this section. We are confident, however, that once a semantics for labeled MSDRT and MSDs as mental states has been spelled out in sufficient detail, adapting that definition to the mental states of this section will be no less straightforward than it it has proved to be for label-free MSDRT.

by content representations, but in such a way that we do not yet know what contents those content representations will give us at a point where we need those contents in order to find out precisely that. It is still necessary to iterate the correctness definition sketched in the present section and the definition of the verification conditions of *Att*-Conditions that builds on it so as to cover mental state describing MSDs whose content specifications may contain *Att*-predications. But that is all right so long as the recursion only concerns the mental state descriptions and not the mental states themselves.

This may look like progress, but how much progress really is it? The obstacle we encountered in Section 3.1 isn't just a technical problem. What lies behind it are deep questions about the semantics of iterated attitude attributions. MSDRT's representations for such attributions may not solve those questions, but it least they give us ways of thinking about them. A non-representational mental state ontology like the one of this section sweeps all this out of sight.

But these questions cannot be kept out of sight forever. One reason for this has to do with a fundamental problem with non-representational approaches to mental content. There may be some merit to the position of the anti-representationalist who argues that in our attempts to understand the mind we should not commit ourselves to detailed assumptions about mental representation for which we have slant evidence if even that; and that there is much we can learn and say about the mind that doesn't require detailed assumptions about mental representation and that we are well-advised to stick to the issues that can be addressed without making such assumptions. But on the other hand, there must be representational formats that the mind uses to store information or which it employs when processing information in certain ways. That some forms of representation must be involved in all of this is, it seems to us close to a tautology. The challenge is to find out what those forms are.

For all we know, this may be a very hard problem, not only because for such a long time because so little progress has been made with it for such a long time, and not for lack of trying, but because the mind makes use of multiple representation formats for the same kinds of information, choosing between them depending on the mental operation for which the information is needed, or transforming representations in one format into representations in another.

If this is right, then the point may come where models of the mind in which there is no room for representation will be recognized to have lost all cognitive plausibility. And that will apply also to the models that are needed in a semantics of attitude attributions. At that point the problem about iterated attributions can be expected to be with us again.

But this is speculation about what is likely to be a quite distant future. For us the sobering conclusion right now is that the semantics of iterated attribution is something of which we do not understand nearly enough.

4.6 A Model-theoretic Account of Attitude Attributions with Empty Names?

In this section we address a problem that has so far fallen between the cracks of our model theory: attributions involving empty names. In Section 2 of Ch. 3 we discussed attitude attributions involving the empty name *Vulcan*. The sentence "Le Verrier assumed that Vulcan was closer to the sun than Mercury", given in Section 2 of Ch. 3 as (3.13), makes an attribution to Le Verrier that has a strong claim to being true in spite of the fact that *Vulcan* has no referent. Two utterances of (3.13), involving users with different views about *Vulcan*, led to the distinct MSDs (3.15) and (3.16), the second of which we repeat here.

$$
\left\{
\begin{array}{l}
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{speaker}(s_H) \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, l_H], \begin{array}{|c|}\hline \\ \hline \text{astronomer}(l_H) \\ \text{Named}(l_H,\ \textit{(Le Verrier)} \\ \hline \end{array}, \mathcal{K}_{l_H}, +real \right\rangle \\[2ex]
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{sun}(l_L H) \\ \text{N'd}(l_H, \textit{(Sun)} \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{N'd}(m_L, \textit{M'ry}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\[2ex]
\left\langle [ENT, v_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_H) \\ \text{Named}(v_H, \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_H}, -\mathbf{real} \right\rangle \\[2ex]
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline t = n \quad t \subseteq s \\
s\!: Att \left(l_H, \left\{
\begin{array}{l}
\left\langle [ENT, m_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\
\left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_L) \\ \text{Named}(s_L, \textit{the Sun}) \\ \hline \end{array}, \mathcal{K}_{s,L}, +real \right\rangle \\
\left\langle [ENT, v_L], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_L) \\ \text{Named}(v_L,\ \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_L}, +real \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline s_1 \\ \hline n \subseteq s_1 \\ s_1\!: \text{Closer-to'}(v_L, s_L, m_L) \\ \hline \end{array} \right\rangle
\end{array}
\right\}, \left\{ \begin{array}{l} < s_L, s_H > \\ < m_L, m_H > \\ \mathbf{LINK}(\mathbf{v_L}, \mathbf{v_H}) \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

The only difference between (3.16), the case where H is of the opinion that *Vulcan* does not refer, and (3.15), the one where H of *Vulcan* as a normally referring name, is in the two parts shown in (3.16) in bold-face. Neither of these is part of the MSD that occurs as argument of *Att*. This MSD is the same in (3.16) and (3.15). By this criterion, the possessors of the mental states portrayed in (3.15) and (3.16) display their attributions to Le Verrier in identical ways. We will assume that it is the MSD arguments of *Att*-predications in our representations that are decisive for the question whether an attitude attribution is correct: the attribution made by a sentence S or discourse D is correct according to thr Logical Form for S or D iff the argument MSD of the relevant *Att*-predication in this Logical Form stands in the right relation to the actual mental state of the attributee.[36]

But what is it for an MSD like the one occurring as *Att* argument in (3.16) to stand in that relation to Le Verrier's actual mental state at the relevant time. And, antecedent to that, how exactly are we to identify Le Verrier's actual mental state at any given time? These will be the central questions of this section.

The first question concerns the mental states of which the MSDs from Logical Forms of attitude attributions are offered as descriptions. Let us go back to the assumption made through most of Section 7, according to which mental states have the form of MSDs. (What follows can be said also on the assumption that mental states are structures of the form (4.30); we leave this as an exercise.) So far we have assumed that all entity representations of mental states properly represent. But what if they don't? According to the model theory we have given nothing much at all – evaluation of such a mental state of MSD of an agent a at any index $<w,t>$ will abort, assuming that the distinguished dref x of a non-denoting ER occurs in the content specification of some propositional attitude in MSD. So the normal basis for assessment of the correctness of an MSD MSD_{Descr} as description of MSD will fail. What we need is an alternative way of assigning values to the content specifications of MSD at indices and on the basis of that some kind of intensional content.

Here is a proposal. Its plausibility hangs on this assumption that our models are rich in that when an agent a is in mental state MSD with one or more empty ERs at some index $<w,t>$ at which an attribution is made to a with a Logical Form in which MSD_{Descr} occurs as argument of Att, then there are other indices $<w',t'>$ at which a is in a state MSD' that is like MSD except that all its ERs properly represent, including the non-representing ones of MSD. For each such

[36]So it is this argument MSD that plays the part of the 'display' of the attribution by the attributor in the sense of the display theory of attribution correctness of (Sainsbury 2018). Whether this is the right interpretation of Sainsbury' display theory is another matter, which can only be settled in discussion with him.

index $<w',t'>$ it will then be possible to assign intensions to the content specifications of both MSD' and MSD_{Descr} and on the basis of that to determine whether MSD_{Descr} is a correct description of MSD'. If MSD_{Descr} is a correct description of MSD' for all such indices $<w',t'>$ by our earlier criteria, then MSD_{Descr} counts as a correct description of a's mental state at $<w,t>$, otherwise not.[37]

Let us see how this proposal pans out for an attribution whose Logical Form is that in (3.16), repeated at the outset if this section. We also repeat on assumption of the scenario: Le Verrier believed at the time at which the attribution says of him that he was in a mental state described by the MSD from its *Att*-predication that the planet for which he had proposed the name *Vulcan* existed, just as the attributor and the interpreter assumed, as witnessed by the +real feature of the ER for Vulcan in the MSD from (3.16)'s *Att*-predication. This means that the content of the belief from Le Verrier's mental state that is attributed to him by (3.16) cannot be evaluated at the index or indices at which the attribution says he was in the state described. But we assume that there are other possible worlds w' in which Le Verrier came up with the same hypothesis for observed aberrations in the movements of Mercury in the context of which he had postulated the existence of a planet he had named *Vulcan* and in which he had been right in that the solar system did contain a planet with the properties he attributed to it.[38] Moreover, there would have to be worlds among these in which there are times at which attributions are made to Le Verrier with (3.16) as their Logical Form. It is in terms of those world-time pairs that the correctness of (3.16) is then to be assessed.

The link between the attributee's command of an empty name on the one hand and the attributor's and interpreter's command of that name on the other – witnessed in (3.16) is witnessed by the item 'LINK(v_L,v_H)' in connection with the name *Vulcan* – takes on a somewhat different complexion in the light of what is proposed in (Kamp 2021*b*) about fictional characters. According to this proposal fictional names can and should be treated as names of fictional characters, where the fictional character denoted by a fictional name N is the network of N-labeled

[37]Note that this new correctness definition can also be applied to cases where MSD does not have any non-representing ERs. This is now a special case where $<w,t>$ is one of the indices at which a is in a state MSD' that is like MSD in the way described (viz. MSD itself). We leave for another occasion the question whether this gives a better correctness definition than the one proposed in Section 3.

[38]It isn't all that easy to argue persuasively for the existence of such worlds, at which this counterfactual situation obtains and everything is in sync. According to both Newtoinian mechanics and its modern successors, together with basic observations about the masses of the sun and Mercury and the distance between them, there could not be a planet with the properties Le Verrier attributed to Vulcan. So eithr physical law or our solar system would have had to be quite different in w from what they are for such a scenario to hold in w. This is an often encountered problem about outlandish possibilities that philosophers are willing to consider.

ERs that members of the language community have and that are connected with each other via their vicarious anchors. (A referring name N that is shared by members of a given speech community also give rise to N-labeled ERs that are connected by such anchors, here the ERs are tied together by the network as well by the referent. But as we have seen in Section 5, the use of a referring name N by the attributor in an attitude attribution to an agent does not require that the N-labeled ER that grounds the attributor's of N must be familiar to the agent (let alone that the agent must have an N-labeled ER for the referent); in those cases it is enough if attributor and agent both have ERs for the given referent. But when there is no referent to license the use of N in an attribution, then agent and attributor must at least have N-labeled ERs denoting the same fictional character. Of course, this disjunction is not an exclusive one. The N-labeled ERs of attributor and attributee can be coreferential in the sense of representing the same real entity and at the same time belong to the same network. And if the attributor knows that, she is justified in her use of N on either or both counts.[39]

With the distinction between +real and -.real ERs comes a suite of further distinction and potential complications. For a flavor of this consider the attributions that our contemporary Mary makes for the benefit of her contemporary Fred that Le Verrier believed that Vulcan was closer to the sun than Mercury. Mary assumes that Fred is like herself in believing that there is no such planet as Vulcan. The Logical Form for the iterated attribution just described, and presented as *oratio recta* in (4.31), involves two -real *Vulcan*-labeled ERs – one as part of the mind of the interpreter H and the other as past of the MSD describing the mind of the outer attributee Fred) – and one +real *Vulcan*-labeled ER as part of the MSD describing the mental state of Le Verrier. The Logical Form is given in (4.32) and (4.33).

(4.31) Fred thinks that Le Verrier believed that Vulcan was closer to the sun than Mercury.

(4.32)

[39] In (Kamp 2021*b*) the notion of a fictional character as the referent of a possibly empty name was limited to fictional names – names of the fictional characters that are found in myths, fairy tales and the works of fiction by individual authors. An implicit assumption there was that the different users all have N-labeled ERs for the fictional name N all of which have the feature -real. But there is no reason to restrict the proposal to just such cases. In particular, there are plenty of names of which some users think that they properly refer while other users do not (as we in fact assumed in the scenario for the use of *Vulcan* that led to H's mental representation in (3.16). In such situations too the name can be used perfectly well in communication, including attributions by those who have -real ERs for N to those with +real ERs for N – the case we have been considering – and attributions by those who have +real ERs for N to those with -real ERs for it.

$$
\left\{
\begin{array}{l}
\left\langle [ENT, fr_H], \begin{array}{|c|}\hline \\ \hline \text{Named}((fr_H, \textit{Fred}) \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[1em]
\left\langle [ENT, l_H], \begin{array}{|c|}\hline \\ \hline \text{astronomer}(l_H) \\ \text{Named}(l_H,\ \textit{(Le Verrier)}) \\ \hline \end{array}, \mathcal{K}_{l_H}, +real \right\rangle \\[1em]
\left\langle [ENT, s_H], \begin{array}{|c|}\hline \\ \hline \text{sun}(l_H) \\ \text{Named}(l_H, \textit{(the Sun)}) \\ \hline \end{array}, \mathcal{K}_{s_H}, +real \right\rangle \\[1em]
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_H) \\ \text{Named}(m_H, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,H}, +real \right\rangle \\[1em]
\left\langle [ENT, v_H], \begin{array}{|c|}\hline \\ \hline \text{Named}(v_H, \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_H}, -real \right\rangle \\[1em]
\left\langle BEL, \begin{array}{|c|}\hline t \quad s \\ \hline t = n \quad t \subseteq s \\
s\text{: } Att \left(fr_H, \left\{ \begin{array}{l}
\left\langle [ENT, m_F], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_F) \\ \text{Named}(m_F, \textit{Mercury}) \\ \hline \end{array}, \mathcal{K}_{m,F}, +real \right\rangle \\[1em]
\left\langle [ENT, s_F], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_F) \\ \text{Named}(s_F, \textit{the Sun}) \\ \hline \end{array}, \mathcal{K}_{s,F}, +real \right\rangle \\[1em]
\left\langle [ENT, v_F], \begin{array}{|c|}\hline \\ \hline \text{Named}(v_F,\ \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_L}, -real \right\rangle \\[1em]
\left\langle BEL, \begin{array}{|c|}\hline t \quad s_1 \\ \hline t \prec n \quad t \subseteq s_1 \\
s_1\text{: } Att \left(L_F, MSD_L, \left\{ \begin{array}{c} < s_L, s_F > \\ < m_L, m_F > \\ LINK(v_L, v_F) \end{array} \right\} \right) \\ \hline \end{array} \right\rangle
\end{array} \right\}, \left\{ \begin{array}{c} < s_L, s_F > \\ < m_L, m_F > \\ LINK(v_L, v_F) \end{array} \right\} \right) \\ \hline
\end{array} \right\rangle
\end{array}
\right\}
$$

(4.33)(MSD_L from (4.32).)

$$\left\{ \begin{array}{l}
\left\langle [ENT, s_L], \begin{array}{|c|}\hline \\ \hline \text{sun}(s_L) \\ \text{Named}(s_L,\textit{(le Soleil})) \\ \hline \end{array}, \mathcal{K}_{s_L}, +real \right\rangle \\[2em]
\left\langle [ENT, m_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(m_L) \\ \text{Named}(m_L,\textit{Mercure}) \\ \hline \end{array}, \mathcal{K}_{m,L}, +real \right\rangle \\[2em]
\left\langle [ENT, v_H], \begin{array}{|c|}\hline \\ \hline \text{planet}(v_H) \\ \text{Named}(v_H, \textit{Vulcan}) \\ \hline \end{array}, \mathcal{K}_{v_H}, +real \right\rangle \\[2em]
\left\langle BEL, \begin{array}{|c|}\hline t' \quad s' \\ \\ \hline \\ t' = n \quad t' \subseteq s' \\ s'\text{: Closer-to'}(v_L,s_L,m_L) \\ \hline \end{array} \right\rangle
\end{array} \right\}$$

We consider it an essential feature for the Logical Forms of attitude attributions involving empty names which of the entity representations in the MDSs offered as descriptions of attributees' mental states have the feature +real and which the feature -real. In the Logical Forms for iterated attributions like the one in (4.31) this question will arise at two or more places, as illustrated in (4.32)/(4.33).

Here we end, for the time being, our proposal for making sense of the question of correctness for attitude attributions to agents who are on the misapprehension that certain names properly refer, although they don't. As it stands it doesn't deal with cases where the attributee doesn't believe of a non-referring name that it does. These cases have to be handled in a quite different way. For a proposal within a MSDRT-based setting see (Kamp 2021*b*).

To conclude this section a brief and preliminary comment to the proposal made in (Sainsbury 2018) for how to characterize correctness of attitude attributions. For Sainsbury correctness is a matter of how the attribution 'displays' the attitudes it attributes. Whether this display is right or wrong is independent from the question whether names occurring in the attribution do or not refer – this is one aspect of the obvious but no less important general principle that correctness is orthogonal to truth of the attributed content in the real world. An attempt to put the proposal for correctness made in Chapter 4, including the one made for attributions containing empty names in the present Section 6, side by side with display theory would naturally involve the assumption that Sainsbury' display can be identified with the MSDs occurring in the Logical Forms advocated here. On this implementation of display theory correctness still has to do with semantics-based logical relations like entailment between the display and what it aims to display, something we believe Sainsbury would not want to advocate. We do not want to prejudge the matter any further here, but just observe that our definition of correctness in terms of entailment or equivalence is independent of the truth of the attributed content; that is a matter of the relation of the attributee's mental state, correctness is a matter of the relationship between that state and the display provided by the attribution.

There is one further caveat here. In Section 3 we made it part of the correctness of attitude attributions that the entities represented by the entity representations attributed to the attributee have certain properties that the attributors themselves attribute to those entities according to the Logical Form of the attribution. (This constraint is imposed by the coreference links $<v,v'>$ in the Links argument of *Att*.) Personal communication with Mark Sainsbury has suggested to me that he would not want to include those constraints in the definition of correctness. It was with that in mind that in the present Section 6 we have not included them in the proposal we have made here. Whether such aspects of attitude attributions, which concern the attributors more directly, should or shouldn't be part of correctness is another matter that still needs sorting out.

4.6.1 Once more: When attributees have two names for the same entity without realizing that they do.

In Section 6 we looked at several examples of attitude attributions to an attributee A who has an N-labeled ER $ER(N)$ and an N'-labeled ER $ER(N')$ which are coreferential but of which he wrongly assumes that they represent different entities. A number of such cases are discussed in the philosophical literature, usually with the tacit assumption that the attributor is aware of all this – i.e. that A has an N-labeled and an N'-labeled ER that represent the same entity, but of which A erroneously assumes that they represent different ones, and, usually by implication, that this is also applies to the interpreter. These were also for the most part the cases we considered in Section 6.

Towards the end of the last section we mentioned Sainsbury's display theory of the correctness of attitude attributions. That was in connection with the use in attitude attributions of empty names like Vulcan and the problem was that new assumptions were needed to guarantee denotations for the content specification of the relevant constituent of mental state of the attributee and of the mental state description provided by the attribution. In Section 6 we looked at several examples in which an attributor S makes an attribution to an attributee A who has an N-labeled ER $ER(N)$ and an N'-labeled ER $ER(N')$ which are coreferential but of which the attributee assumes that they represent different entities. S is aware of all this, wants to use either N or N' in her attribution and must choose the intuitively right one of these two names. Here the problem is not a matter of getting proper denotations, but of finding criteria that distinguish between the two name options, although the choice makes no difference to the denotation.

As just described, the correctness problem for these attributions isn't a denotation problem.[40] So it might be objected that Section 7 isn't the right place to address this problem. It is only at this point, however, that the denotations for content specifications in models have been defined. So it is only now that we can see plainly that given our intensional approach, the problem cannot be solved at the level of denotation. So what little we have to say about this problem might as well be said right here.

Let us focus once more on cases involving *Hesperus* and *Phosphorus* and consider the one in which an attributor S today makes a belief attribution to some person A from ancient Greece, assuming that A has a *Hesperus*-labeled and a *Hesperus*-labeled ER and the false belief that these represent distinct entities, and furthermore that the attributee believes that the entity represented by his *Hesperus*-labeled ER can be seen only for some time after sunset and that the entity represented by his *Phosphorus*-labeled ER can be

[40] At least, it isn't that unless one sees the intuitions discussed in Section 6 are seen as indicating that the intensions our model theory offers as denotations for the contents of propositional attitude components of mental states and of the content specifications of attributions aren't fine-grained enough. But as noted early on in the present Section 7, that is a major question we cannot go into in this document.

seen only for some time before dawn. And let us assume for simplicity also that S takes it for granted that her interlocutor H makes the same assumptions about the attributee. What more can we say, given these assumptions about attributions that S can make to A for the benefit of H beyond what was said about these in Section 6?

Intuitively, the choice that S is facing between *Hesperus* and *Phosphorus* is to decided on the basis of which one of A's *Hesperus*-labeled and a *Phosphorus*-labeled ERs is involved in the belief that S wants to attribute to him. According to what has been said about the use of names in the complement clauses of attitude attributions in Section 5, the use of N in an attitude attribution to A only requires that S herself have an N-labeled ER for an entity $\mathbf{d}$ and correctly assumes that the content of the attitude she wants to attribute to A involves an ER of A's that also represents $\mathbf{d}$; there is no need for A's ER to also have N as a label, coreference of the two ERs suffices. To extend this principle to the case before us we need to say a little more about S's own ER for the planet Venus, which she takes to be coreferential with both the *Hesperus*-labeled ER and the *Phosphorus*-labeled ER she assumes A to have. Let us assume that S's ER for Venus to have all three names *Hesperus*, *Phosphorus* and *Venus* as labels. So in fact S has the choice from three names when she wants to attribute to A an attitude about Venus.

Suppose that S wants to attribute to A the belief and desire that A would have expressed with the words (translated into English as) (4.34.a). In this case it seems clear that (4.34.b) is a correct attribution and that (4.34.c) is not. In fact most versions that can be obtained from (4.34.a) by replacing one or more of its name occurrences with one of the other two names from the set $\{$*Hesperus*, *Phosphorus*, *Venus*$\}$ doesn't seem to qualify as a correct report of what A himself would express as in (4.34.a).

(4.34)a. I never saw Phosphorus. I would like to see Phosphorus at least once. Hesperus I have seen many times.

b. A thought that he had never seen Phosphorus and that he wanted to see Phosphorus at least once. He thought he had seen Hesperus many times.

c. A thought that he had never seen Hesperus and that he wanted to see Hesperus at least once. He thought he had seen Phosphorus many times.

The reason why it seems intuitively clear to us that if A would have expressed his beliefs and want as in (4.34.a), then (4.34.b) would have been a correct attribution and that (4.34.c) would not have been is that the relevant part of A's mental state must have been as shown in (4.35).

$$
(4.35)\left\{
\begin{array}{l}
\left\langle [ENT,\eta_A], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\eta) \\ \text{N'd}(\eta_A,\ \textit{Hesperus}) \\ \\ \hline \end{array}, \mathcal{K}_{\eta_A}, +real \right\rangle \\
\left\langle [ENT,\phi_A], \begin{array}{|c|}\hline \\ \hline \text{cel.body}(\phi_A) \\ \text{N'd}(\phi,\textit{Phosphorus}) \\ \\ \hline \end{array}, \mathcal{K}_{\phi_A}, +real \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \eta_A \neq \phi_A \\ \\ \hline\end{array} \right\rangle
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \neg \begin{array}{|c|}\hline e \\ \hline e \prec n \\ e\text{: see'}(i,\phi_A) \\ \hline\end{array} \\ \hline \end{array} \right\rangle
\left\langle DES, \begin{array}{|c|}\hline e' \\ \hline n \prec e' \\ e'\text{: see'}(i,\phi_A) \\ \hline \end{array} \right\rangle \\
\left\langle BEL, \begin{array}{|c|}\hline \\ \hline \begin{array}{|c|}\hline t \\ \hline t\prec n \\ \hline\end{array} \left\langle\begin{array}{c} MANY \\ t\end{array}\right\rangle \begin{array}{|c|}\hline e'' \\ \hline e''\subseteq t \\ e''\text{: see'}(i,\eta_A) \\ \hline\end{array} \\ \hline\end{array} \right\rangle
\end{array}
\right\}
$$

For an attribution to A of the beliefs and desire of A shown in (4.35) to be correct the Logical Form for this attribution must contain an MSD in 3rd argument position of *Att* that can be suitably embedded into (4.35) by functions TRANS and TARGET. And that requires that this MSD have ERs for (intuitively speaking) Hesperus and Phosphorus, such that the distinguished dref for the Phosphorus ER occurs as second argument of 'see" in the first belief and the desire constituent of the MSD and the distinguished dref for the Hesperus ER occurs as second argument of 'see" in its second belief constituent. These requirements can be expected to be satisfied by the Logical Form for (4.34.a) but not by the Logical Form for (4.34.b). And if that is so, we have an account of our intuitive judgments that (4.34.a) describes A's mental state correctly and that (4.34.b) does not.

But can we be sure that the Logical Forms of (4.34.a) and (4.34.b) must be such that the first satisfies these requirements and the second does not? That, unfortunately, can not be concluded from what has been stipulated so far. The reason is that in view of what we have said about the use of names in the complement clauses of attitude attributions it is enough when the name used by the attributor labels an ER of hers that is coreferential with the ER that she attributes to the attributee by virtue of using the name. In the case we are discussing the *Hesperus*-labeled and *Phosphorus*-labeled ERs that A has are both coreferential with the S's trice labeled ER for Venus. So according to the coreference requirement for the use of names in complement clauses S could use any one of the labels *Venus*, *Hesperus* and *Phosphorus* to relate her attribution either to A's *Hesperus*-labeled ER or to her *Phosphorus*-labeled ER. That allows S a good deal of choice in cases like the one we are considering, and much more than seems right.[41]

What constraints can be added to rule out (4.34.b) as a correct attribution to A (and many of the other variants of (4.34.a) as well), while licensing (4.34.a) as correct? As things stand, we do not see our way through to a fully general constraint, and we are not sure that a reasonably concise general constraint can be found. For the case before us, where S has a clear idea about the entity representation of her attributee A, the intuitively right constraint is that S should point to the N-labeled ER of A by using N – thus *Hesperus* should be used to point to A's *Hesperus*-labeled ER (but not *Phosphorus* or *Venus*), and likewise for A's *Phosphorus*-labeled ER. But while this constraint seems right for the cases where the attributor has such a thorough picture of the attributee's state of mind, it does not cover cases where the picture isn't as detailed. More work is needed to see what additional constraints may be needed for these other cases.

How does this discussion tie up with Sainsbury's display theory? Once again,

[41] A curious case is that where S uses *Venus* throughout her attribution. That makes the attribution sound like attributing contradictory beliefs to A, on the one hand the belief that he, A, never saw Venus and on the other hand the belief that he saw Venus many times. That makes the attribution sound weird, and it would make a weird, and misleading impression on a recipient who didn't share Ss assumptions about A. The effect of the attribution on the recipient – the Logical Form that he will construct for it – is an aspect of the problem we are discussing that we are not saying much about in the present section. But we trust that enough has been said about the communication-theoretic approach to attitude attribution to enable the readers to fill in these gaps for themselves.

this is a matter that it is difficult for us to say much about without further consultation. As preliminaries just the following two points. (i) The display provided by the complement clauses of the attributions is sensitive to the choice of names (from a set of two or more coreferential names). (ii) The question what counts as a correct display may have something to do with the extent of the knowledge that the attributor has of the mental state of the attributee and perhaps also with the assumptions that the interpreter is in a position to make about the speaker's knowledge. But note that once again these uncertainties have got to do with the rules for converting the attributions as sentences or sentence sequences into their Logical Forms. As far as we can see, for the Logical Forms themselves the correctness definition can remain as given.

MORE TO COME

Currently there are two more Chapters under construction. The last one presents applications of the framework introduced in this document to problems of individual and joint deliberation, planning and action, with an emphasis on discussions of the work of Bratman and Ludvig. The chapter preceding this last one lays some of the formal foundations for the proposals that will be made in the last chapter by laying out the needed parts of DRT's treatment of plurals and of collectivity as a mental category. Especially the last chapter so far currently exists only in a very finished form. Since it will not be a topic for current purposes, it seemed better to leave it out here, and with it the chapter on plurality that is leading up to it.

Bibliography

Abney, S. (1987), The English Noun Phrase in its Sentential Aspect, PhD thesis, MIT.

Abusch, D. (1997), 'Sequence of tense and temporal de re', *Linguistics and Philosophy* pp. 1–50.

Asher, N. (1986), 'Belief in discourse representation theory', *Journal of Philosophical Logic* **15**, 127–189.

Asher, N. (1993), *Reference to Abstract Objects in Discourse*, Kluwer, Dordrecht, Dordrecht.

Barker, C. (1995), *Possessive Descriptions*, University of Chicago Press.

Bary, C. & Altshuler, D. (n.d.), Double access, *in* 'Proceedings of Sinn und Bedeutung'.

Beaver, D. (1997), Presupposition, *in* J. van Benthem & A. ter Meulen, eds, 'Handbook of Logic and Language', Elsevier, Amsterdam, pp. 939–1008.
URL: *citeseer.nj.nec.com/beaver96presupposition.html*

Beaver, D. (2001), *Presupposition and Assertion in Dynamic Semantics*, CSLI.

Bratman, M. (1987), *Intention, Plans and Practical Reasoning*, CSLI Publications.

Bratman, M. (2014), *Shared Agency: A Planning Theory of Acting Together*, OUP.

Champollion, L. (2014), 'The interaction of compositional semantics and event semantics', *Linguistics and Philosophy* **38**, 3166.

Cocchiarella, N. (1965), Tense and Modal Logic, PhD thesis, UCLA.

de Swart, G. & Molendijk, A. (1999), 'Negation and the temporal structure of narrative discourse', *Journal of Semantics* **16**, 1–42.

Degen, K. & Tonhauser, J. (2022), 'Are there factive predicates? an empirical investigation', *Language* **98(3)**, 552–591.

Dowty, D. (1979), *Word Meaning and Montague Grammar*, Reidel, Dordrecht, Dordrecht.

Enç, M. (1981), Tense without Scope: An Analysis of Nouns as Indexicals., PhD thesis, University of Wisconsin at Madison.

Enç, M. (1986), 'Towards a referential analysis of temporal expressions', *Linguistics and Philosophy* **9**, 1–44.

Farkas, D. (2002), 'Specificity distinctions', *Journal of Semantics* **19**, 1–31.

Frank, A. (1997), Context Dependence in Modal Constructions, PhD thesis, IMS, University of Stuttgart.

Frank, A. & Kamp, H. (1997), On context dependence in modal constructions, *in* 'Proceedings of SALT VII', CLC Publications and Cornell University.

Gallin, D. (1975), *Intensional and Higher Order Logic; with Applications to Montague Semantics*, North Holland, Amsterdam.

Gärdenfors, P. (1988), *Knowledge in Flux: Modelling the Dynamics of Epistemic States*, MIT Press, Cambridge Mass.

Gärdenfors, P. (1992), *Belief Revision*, Cambridge University Press.

Groenendijk, J. & Stokhof, M. (1990), Dynamic montague grammar, *in* L. Kalman & L. Polos, eds, 'Papers from the Second Symposium on Logic and Language', Akademiai Kiadoo, Budapest, pp. 3–48.

Groenendijk, J. & Stokhof, M. (1991), 'Dynamic predicate logic', *Linguistics and Philosophy* **14**, 39–100.

Groenendijk, J., Stokhof, M. & Veltman, F. (1996*a*), Coreference and modality, *in* S. Lappin, ed., 'Handbook of Contemporary Semantic Theory', Blackswell, Oxford, pp. 179–216.

Groenendijk, J., Stokhof, M. & Veltman, F. (1996*b*), This might be it, *in* D. Westerstahl & J. Seligman, eds, 'Language, Logic and Computation: The 1994 Moraga Proceedings', CSLI, Stanford, pp. 255–270.

Groenendijk, J., Stokhof, M. & Veltman, F. (2004), Coreference and modality in multi-speaker discourse, *in* H. Kamp & B. Partee, eds, 'Context Dependence in the Analysis of Linguistic Meaning', Elsevier, pp. 195–216.

Hansson, S. (1999), *A Textbook of Belief Dynamics. Theory Change and Database Updating.*, Kluwer.

Heim, I. (1982,1988), *The Semantics of Definite and Indefinite Noun Phrases*, Garland Press.

Heim, I. (1983), 'On the projection problem for presuppositions', *Proceedings of the West Coast Conference on Formal Linguistics* **2**, 223–248.

Higginbotham, J. & Schein, B. (1989), Plurals, *in* 'North East Linguistics Society (NELS), Vol. 19'.

Kamp, H. (1981*a*), 'Evénements, représentations discursives et référence temporelle', *Langages* **64**, 39–64. For the original English ms. from which the paper in *Langages* is the French translation, see *Semantics and Pragmatics*, Vol 10 (2017.

Kamp, H. (1981*b*), A theory of truth and semantic representation, *in* J. Groenendijk et al., eds, 'Formal Methods in the Study of Language', Mathematisch Centrum, Amsterdam. Reprinted in: K. von Heusinger and A. Ter Meulen, 'Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp'. Brill, 2013.

Kamp, H. (1999), Intentions, plans and their execution: Turning objects of thought into entities of the external world. ms. Stuttgart.

Kamp, H. (2003), Einstellungszustände und einstellungszuschreibungen in der diskursrepräsentationstheorie, *in* U. Haas-Spohn, ed., 'Intentionalität zwischen Subjektivität und Weltbezug', mentis, Paderborn, pp. 209–289.

Kamp, H. (2015), 'Using proper names as intermediaries between labelled entity representations', *Erkenntnis* **80**, 263–312.

Kamp, H. (2017), 'Events, discourse representations and temporal reference', *Semantics and Pragmatics* **10(2)**, 1–45. A French translation of this article appeared in *L*angages, 1981.

Kamp, H. (2019), Tense and aspect in discourse representation theory, *in* R. Truswell, ed., 'The Oxford Handbook of Event Structure', Oxford University Press.

Kamp, H. (2021*a*), Notes semantics: Drs construction; university of texas. Unpublished notes for a graduate course in Linguistics offered at the University of Texas, Austin.

Kamp, H. (2021*b*), Sharing real and fictional reference, *in* E. Maier & A. Stokke, eds, 'The Language of Fiction', Oxford University Press.

Kamp, H. (2022*a*), 'The links of causal chains', *Theoria* **88(2)**, 296–325. Special Issue on 'Naming snd Necessity'.

Kamp, H. (2022*b*), Referential and attributive descriptions, *in* D. Altshuler, ed., 'Linguistics Meets Philosophy', Cambridge University Press, pp. 72 – 98.

Kamp, H. & Bende-Farkas, A. (2019), 'Epistemic specificity from a communication-theoretic perspective', *Journal of Semantics* **36 (1)**, 1 – 51.

Kamp, H. & Reyle, U. (1993), *From Discourse to Logic*, Kluwer.

Kamp, H. & Reyle, U. (1996), 'A calculus for first order discourse representation structures', *Journal for Logic, Language and Information* **5**, 297–348.

Kamp, H., van Genabith, J. & Reyle, U. (2011), Discourse representation theory: An updated survey, *in* D. Gabbay, ed., 'Handbook of Philosophical Logic, Vol. XV', Elsevier, pp. 125 – 394.

Karttunen, L. (1974), 'Presupposition and linguistic context', *Theoretical Linguistics* pp. 181–194.

Klein, E. (1981), Defensible descriptions, *in* F. Heny, ed., 'Ambiguities in Intensional Contexts', Reidel, Dordrecht, pp. 83–102.

Krifka, M. (1989), Nominal reference, temporal constitution and quantification in event semantics, *in* R. Bartsch, J. van Benthem & P. van Emde-Boas, eds, 'Semantics and Contextual Expressions', Foris, Dordrecht, pp. 75–115.

Krifka, M. (1998), The origins of telicity, *in* S. Rothstein, ed., 'Events and Grammar', Kluwer.

Kripke, S. (1979), A puzzle about belief, *in* A. Margalit, ed., 'Meaning and Use', Reidel, Dordrecht.

Landman, F. (1989*a*), ''Groups', Part I, II', *Linguistics and Philosophy* **12**, 559–605, 723–744.

Landman, F. (1989*b*), 'The progressive', *Natural Language Semantics* **1**, 1–32.

Landman, F. (2000), *Events and plurality: the Jerusalem lectures*, Kluwer, Dordrecht.

Link, G. (1983), The logical analysis of plurals and mass terms: A lattice-theoretical approach, *in* R. Bäuerle, C. Schwarze & A. von Stechow, eds, 'Meaning, Use and Interpretation of Language', Walter de Gruyter, Berlin, pp. 303–323.

Ludwig, K. (2016), *From Individual to Plural Agency*, OUP.

Maier, E. (2016), Why my i is your you: On the communication of de se attitudes, *in* M. Garca-Carpintero & S. Torre, eds, 'About Oneself: De Se Thought and Communication', OUP.

Maier, E. & Stokke, A. (2020), *The Language of Fiction*, OUP, Oxford, UK.

Montague, R. (1970), 'Universal grammar', *Theoria* **36**, 373–398. Reprinted in: R. Thomason (ed), Formal philosophy; selected papers of Richard Montague, Yale University Press, 1974.

Musan, R. (1999), 'Temporal interpretation and information-status of noun phrases', *Linguistics and Philosophy* **22**, 355–401.

Ogihara, T. (1992), Temporal Reference in English and Japanese, PhD thesis, University of Texas at Austin. Distributed by Indiana University Linguistics Club.

Ogihara, T. (1994), Adverbs of quantification and sequence-of-tense phenomena, *in* 'Proceedings from Semantics and Linguistic Theory (SALT4)', CLC Publications, Cornell University, pp. 251–267.

Ogihara, T. (1999), Double-access sentences generalized, *in* 'Proceedings from Semantics and Linguistic Theory IX (SALT9)', CLC Publications, Cornell University.

Ogihara, T. & Sharvit, Y. (2015), Embedded tenses, *in* R. Binnick, ed., 'Handbook of Tense and Aspect', Oxford University Press, Oxford, U.K., pp. 638–668.

Partee, B. (1973), 'Some structural analogies between tenses and pronouns in english', *Journal of Philosophy* **70**, 601–9.

Partee, B. (1984), 'Nominal and temporal anaphora', *Linguistics and Philosophy* **7**, 243–286.

Partee, B. H. & Borschev, V. (2003), Genitives, relational nouns, and argument-modifier ambiguity, *in* E. Lang, C. Maienborn & C. Fabricius-Hansen, eds, 'Modifying Adjuncts', Mouton de Gruyter.

Prior, A. (1967), *Past, Present and Future*, Clarendon Press, Oxford.

Sainsbury, R. (2005), *Reference without Reference*, Oxford University Press, Oxford, United Kingdom.

Sainsbury, R. (2018), *Thinking about Things*, Oxford University Press, Oxford, United Kingdom.

Sainsbury, R. & Tye, M. (2012), *Seven Puzzles of Thought and How to Solve Them: an Originalist Theory of Concepts*, Oxford University Press, Oxford, United Kingdom.

Scha, R. (1981), Distributive, collective and cumulative quantification, *in* J. e. a. Groenendijk, ed., 'Formal Methods in the Study of Language', Mathematisch Centrum, Amsterdam, pp. 483–512.

Schein, B. (1993), *Plurals and Events*, MIT Press, Cambridge, Mass.

Stanley, J. & Szabo, Z. (2000), 'On quantifier domain restriction', *Mind and Language* **15**, 219–261.

Stowell, T. (2007), 'The syntactic expression of tense', *Lingua* **117(2)**, 437–463.

Tonhauser, J. (2002), A dynamic semantic account of the temporal interpretation of noun phrases, *in* B. Jackson, ed., 'Proceedings of Semantics and Linguistic Theory (SALT) XII', Cornell Linguistics Circle Publications, pp. 286–305.

Van Der Sandt, R. (1992), 'Presupposition projection as anaphora resolution', *Journal of Semantics* **9**, 333–377. Special Issue: Presupposition, Part 2.