The Links of Causal Chains

Hans Kamp

January 3, 2022

1 Orientation

This paper is about just one part of Kripke's many lasting contributions to logic and philosophy. But it is about a central part, about how names work in language and thought, about their uses in overt and silent speech and writing – and, just as importantly, about how they don't work, but had been assumed to work by practically everybody before *Naming and Necessity* turned the world of logic, language, metaphysics and epistemology right side up.

I was part of that world, and my world too was made anew by *Naming and Necessity*. And I too could witness, from that point onwards, the snowballing literature that came in its wake. Over the years that literature has grown to a size where it is hard to keep an overview and I am well aware that I have much less of an overview than many others. But in the part of it that I have seen the predominant focus has been on the logical and semantic implications of what *Naming and Necessity* has to say about proper names: that names are 'rigid designators' and exactly what that entails for the meaning and use of sentences containing names and for the general semantics and logic of languages containing such sentences.

What *Naming and Necessity* has to say about proper names, however, also has an important social dimension. It is a familiar observation that the use of language has a double function. We use language to tell others things that we want them to know, or to get them to do certain things and so on. But the utterances that competent speakers produce for these various purposes also serve as displays of how language *is* to be used, as samples of how it works. This second aspect is especially plain and prominent when utterance recipients are not fully competent speakers, schoolchildren with a foreign background, for example, who struggle to learn the local language from the

lessons they attend, while trying to figure out at the same time what they are being told about the subjects they are taught, history, geography, mathematics, biology and so on. But utterances by competent speakers do not only play their role as samples of proper use when the recipients are not yet fully competent users of the language (according to the informal and not very well defined criteria for linguistics that people willy-nilly go by). They can serve effectively as samples of language use also when their recipients are what we consider competent. There just is no practically applicable notion of complete language competence, and certainly not for human languages today, with their steadily growing special vocabularies to keep apace with cultural diversification and scientific progress. For pretty much all of us language acquisition never stops.

That utterances serve as samples of how their language functions is a tired triviality, which bears no repeating unless followed by something more specific. More specific, and of genuine linguistic and philosophical interest, are the specific mechanisms that make it possible for utterances to serve their recipients as examples of proper language use while at the same time getting the messages across that the speaker or author wants to communicate. Such mechanisms are of interest in their own right and merit closer investigation on that account. But their investigation can also gain us a better understanding of the essential intersubjective dimension of human languages, as tightly structured practices shared by speech communities, which can change in the course of time, but which nevertheless have the remarkable degree of stability without which verbal communication would be impossible. It is because of their stability that languages can be fruitfully studied as abstract, user-independent systems in the spirit of for instance (Morris 1946), (Carnap 1947) or (Montague 1970*b*) and why so much progress has been made in the past fifty years by treating them that way. The mechanisms I have in mind have the potential to explain the stability of language, while accounting at the same time for their capacity to change.

For the most part serious investigation of such mechanisms has been getting under way only in recent years.[1] But one such mechanism, the causal theory theory of names that is outlined in *Naming and Necessity*, has been known to us for more than half a century. Recall the central ideas of the causal chain theory: Proper names do not refer via unique satisfaction of some descriptive content, but because of how they originate, by some form of 'baptismal' ceremony, and then spread through the language community. That is, a name N

[1]For some remarks on a couple of recent developments in this direction see Section 5.

starts out when one or more persons introduce it as 'label' for some entity **r**, thus providing those who are in on the act with a convenient means to refer to **r**. From this initial core of N-users the command of N can then be passed to others: a person S who is in command of N as name of **r** can pass her command to some other speaker H who did not yet have it, either by simply using N in utterances that say something about its referent **r**, or by introducing H explicitly into the use of N.[2] Either way H can pick up the name from what he is told, or hears or reads, and then start using N himself to refer to **r**.

It is through such transfers that members of the speech community at large are inducted into the sub-community of those who are in command of N as name for **r**. In other words, at any time after the introduction of N as a name for **r** this sub-community is composed of (i) members of the 'baptismal core', who were a party to the introduction of *Naming and Necessity* as name of **r** (assuming that any of those are still alive); and (ii) speakers who are connected to one or more people in the core by some 'causal chain' of individual transfers of the command of N. Because of these transfers the sub-community changes all the time and with it, indirectly, the speech community as a whole. But at the same time the transfer mechanism guarantees the stability that matters: the reference of N is in principle always preserved.

In this paper I want to look more closely at the transfer mechanism that is responsible for the links of causal chains. But for what I want to say we need a formal framework. The one I will be using is MSDRT (short for 'Mental State Discourse Representation Theory'), an extension of DRT ('Discourse Representation Theory'). Since I don't assume that readers of the present issue of Theoria are familiar with either framework, I will provide introductions to both DRT and MSDRT, stripping them down to the minimum that I will need for the remarks on causal chain links that follow. Section 2 will be devoted to DRT, Section 3 to MSDRT. Section 4 applies MSDRT to causal chains and their links. The central notion developed in that section will be that of a *network of linked Entity Representations*. Causal chains will be branches within such networks. Some speculative remarks on the generalization of the approach of Section 4 will follow in the concluding Section 5.

[2]For discussion of different ways in which one speaker can introduce another one to the use of a name see (Kamp 2015).

2 DRT

The original motivation for DRT were certain problems in the semantics of tense and aspect. A proper account of the temporal and aspectual properties of tenses must have something to say about how certain tenses relate what is described in the sentences containing them to what is described in earlier sentences.[3] But DRT owed its early visibility to what it has to say about the so-called 'donkey problem', a problem from the medieval scholastic literature that had gained prominence with 20th century philosophers and linguists through its discussion in (Geach 1962). One example of the problem – only one is needed here – is the pronoun *it* in the consequent of the conditional sentence (1), which is anaphoric to the indefinite NP *a donkey* in the conditional's antecedent.

(1) If Bill owns a donkey, he beats it.

Geach observes that (1) has truth conditions roughly captured by the paraphrase that Bill beats every donkey that he owns. Sentences like (1) present a difficulty for truth-conditional natural language semantics along the lines of Montague[4], the widely accepted standard for formal semantics in the late seventies. More precisely, if we follow Geach in assuming that (1) has the 'universal' truth conditions given by the paraphrase above, then such semantics is confronted with a dilemma: either the truth conditions come out wrong or the pronoun becomes a 'dangling' variable and the sentence fails to get a proper interpretation altogether.[5]

The account that DRT offers for sentences like (1) is based on the following assumptions. Its point of departure is that the donkey pronoun problem

[3] The first tenses that led to this conclusion were the French past tense forms Passé Simple and Imparfait. These are both past tenses, but they differ in that the Passé Simple tends to express temporal progression: an event described by a Passé is typically understood to have occurred after the event described by a preceding sentence, usually the immediately preceding one. In contrast, the Imparfait normally conveys simultaneity between the state of affairs described by its sentence and the event or state described by a preceding sentence. A similar contrast to that between Passé Simple and Imparfait can be observed for the Simple Past and the Past Progressive of English; compare for instance the two-sentence discourses "John turned around. The man behind him pulled his gun." and "John turned around. The man behind him was pulling his gun.' For discussion see e.g. (Kamp, van Genabith & Reyle 2011).

[4] see in particular (Montague 1970*a*), (Montague 1970*b*) and (Montague 1970*a*).

[5] There has been much debate over the question if (1) really does have the truth conditions Geach assumed. For present purposes this matters little, since other readings that have been suggested present similar difficulties for a Montagovian semantics. And in any case there remains the question how to deal with 'donkey discourses', such as (2) below.

of sentences like (1) is at bottom a discourse phenomenon, much like the interpretation of tense forms in multi-sentence discourses of the sort described in footnote 3. (And by the way, the tense forms mentioned there also manifest their anaphoric properties in conditionals and other complex sentences.) To deal with Geach's problem it is therefore recommendable to start with a *donkey discourses*, such as (2).

(2) Bill owns a donkey. He beats it.

According to DRT the interpretation of this sentence takes the form of first constructing a semantic representation/logical form[6] for the first sentence and then making use of that representation in constructing a semantic representation for the second sentence. The logical form for the two sentence discourse in (2) is then obtained by merging the two representations for its two sentences.

Formally the representation of the first sentence is that in (3).

(3)

b y
Bill'(b) donkey'(y) own'(b,y)

(3) is an example of a so-called *Discourse Representation Structure* or *DRS*. DRSs are DRT's logical forms. A DRS consists of two parts, (i) its *Universe* (in (3) this is the part above the horizontal bar in the middle), which is a set of *discourse referents* (or *drefs*) and (ii) its *Condition Set* (the part below the bar), a set of *DRS Conditions*. Discourse referents represent entities, DRS Conditions are typically open formulas which may contain drefs occurring in the Universe of the DRS (as well as, often, drefs not occurring there) in argument positions of predicates. Just as formulas of Predicate Logic can be regarded as formulas belonging to particular languages of Predicate Logic (determined by their non-logical vocabularies), so DRSs can be regarded as belonging to particular DRS languages. (For instance, a DRS language that includes (3) among its formulas will have the 1-place predicates *Bill'*, *donkey'* and the 2-place predicate *own'* in its vocabulary.

[6]My use of the term 'logical form' in this paper is in essence that familiar from Philosophical Logic, where formulas of Predicate Logic are used to describe the truth conditions of sentences from natural language. No distinction is made in this paper between 'logical form' and 'semantic representation'.

DRT comes with an explicit model-theoretic semantics for its DRS languages. The models for these languages are like models for Predicate Logic. For instance, a model M for a DRS language that contains (3) as one of its 'formulas' will assign to *Bill'* and *donkey'* subsets of its Universe U_M, and to *own'* a 2-place relation-in-extension, i.e. a set of ordered pairs of members of U_M. (On the assumption that a name is the name of one and no more than one individual, the extension of *Bill'* in M will be a singleton. If you do not like this as an account of proper names, don't worry. The point of this paper is to do better.)

In the model-theoretic semantics for DRS languages the drefs in the Universe of DRSs are interpreted existentially. This existential interpretation results from the general principle that a DRS $K = <U,C>$ is true in a model M iff there exists a *verifying embedding* of K in M – a function f from U into the Universe U_M of M such that any Condition γ in C is true in M under the assignment that f provides for the drefs in γ. (Example: (3) is true in a model M if there exists a function f that maps b to an individual **b** in U_M and y to an individual **y** such that **b** is in the extension in M of *Bill'*, **y** is in the extension in M of *donkey'* and the pair <**b,y**> belongs to the extension in M of *own'*.

The construction of a DRS for the second sentence of (2) makes use of (3); (3) serves as context for its interpretation (just as the first sentence of the two sentence discourses in footnote **??** provides the context for the interpretation of the sentence following it). The use that the DRS construction for the second sentence of (2) makes of (3) is determined by an interpretation rule for anaphoric pronouns, according to which personal pronouns can be construed as anaphoric to drefs in the Universe of the context DRS. Using this rule it is possible to construct (4.a) as representation of the second sentence, in which the dref x, introduced to represent the pronoun *he*, is identified with its 'anaphoric antecedent' b in the Universe of (3) and the dref z, introduced to represent the pronoun *it*, is identified with y. After its completion, (4.a) can be merged with the representation (3) of the first sentence. The result, shown in (4.b), is the Logical Form for the two sentences of (2) together.

(4) a.

x	z
$x = b$	$z = y$
beat'(x,z)	

b.

b y x z
Bill'(b) donkey'(y) $x = b$ $z = y$ own'(b,y) beat'(x,z)

To deal with donkey sentences like (1) conditionals have to be treated as built from (i) an antecedent which describes a situation (or a set of possible worlds) and (ii) a consequent, which gives additional information about the situation (or the worlds in the set) described by the antecedent. In this case the semantic representation (i.e. the DRS) of the antecedent provides a context for the interpretation of the consequent, in the same way that the DRS for the first sentence of (2) provides a context for the interpretation of the second sentence of (2). Note well that to make this explicit three further stipulations are needed:

(i) The DRS for a conditional like (2) must be a representation in which the DRSs for antecedent and consequent are connected by an operator that represents their conditional relationship. DRT uses the sign '⇒' for this purpose; it forms DRS Conditions $K \Rightarrow K'$ from DRSs K and K', their antecedents and consequents. So the DRS for (2) takes the form shown in (5), with empty Universe and a Condition Set whose only member is a conditional DRS Condition of the form $K \Rightarrow K'$:

(5)

b y		x z
Bill'(b) donkey'(y) own'(b,y)	⇒	$x = b$ $z = y$ beat'(b,z)

(ii) To construct the DRS in (5) for the sentence in (1) a rule is needed that converts the part of (1) that expresses conditionality – i.e. the combination of *if* in the subordinate clause of (2) and the comma that separates it from the main clause – into the operator ⇒ that connects the DRS for the *if*-clause with the DRS for the main clause. Furthermore, it must be stipulated that the DRS which stands to the left of ⇒ can serve as interpretation context for the DRS on the right. And finally, (iii) the definition of truth for DRSs in models must be extended with the following clause for DRS Conditions of the form $K \Rightarrow K'$:

(6) An embedding function f *verifies* DRS Condition $K \Rightarrow K'$ *in* model M iff for every extension g of f to the Universe of K[7] that verifies K there exists an extension h of g to the Universe of K' which verifies K'.[8]

DRT contends that these stipulations are natural; they capture the intuitive semantics of the words and grammatical constructions they apply to, and do this in a conceptually transparent manner. In fact, most of the work that goes into DRT-based treatments of natural language fragments consists in finding and correctly stating the DRS construction rules for the different words and grammatical constructions of the fragment; in this regard DRT is no different from other approaches to the semantics of natural language that conform to the standards of precision and explicitness set by the work of Montague.

Some DRS construction rules relate the DRS under construction to the context DRS that has been constructed from the preceding part of the discourse. (Our only example of such a rule so far is the one for anaphoric pronouns that we needed to construct the DRSs for (1) and (2); but for present purposes this example is all we need.) It is rules of this kind that make DRT into a genuine theory of *discourse semantics* and not just a semantic theory for single sentences in isolation. But for this very reason they present a special challenge: The structures they relate to each other – the partly constructed representation for the current sentence and the already constructed representation that serves as context – must have forms that make it possible to state these rules in such a way that they produce the right results whenever they are applied. It was the attempt to meet this challenge that led to the representation format of DRSs. And it was thought to be an appealing feature of DRT that it had succeeded to meet this challenge by adopting a representation format for its DRSs that enables them to do both things at

[7] i.e. a function g that coincides with f on the Domain of f and whose domain is the extension of the Domain of f with the Universe of K

[8] Note that the occurrences of the drefs b and y in the consequent DRS of (5) are 'free' in this DRS in that these drefs do not occur in its Universe. When clause (6) is applied to the $\Rightarrow$-Condition of (5) as part of applying the verification definition for DRSs in models to DRS (5), then these drefs are bound 'from the outside': the functions g spoken of in (6) already assign values to b and y, which are then passed on to their extensions h that must verify the consequent DRS. There are also DRSs K with occurrences of drefs that are nowhere bound within K, in that they occur neither in the Universe of K itself no in the Universe of any of its sub-DRSs. Such DRSs are like open formulas of Predicate Logic in that they can be evaluated for truth or falsity only relative to an assignment – a function from drefs to entities in the model – that is determined independently (e.g. by the context in which the content represented by K is being used.)

the same time: correctly capture the content of the bits of language for which they are constructed as logical forms and serve as discourse contexts for the construction of DRSs for further bits.

It is tempting to see this feature of DRT – that the same structures serve as content representations for what has been interpreted already and as contexts for the interpretation of what comes next – as evidence that DRT captures some important aspects of how the information that human interpreters obtain from what they read or hear is mentally represented. Whether this evidence is to be taken seriously, or how seriously, has been a bone of contention from the time when DRT became accessible to the public. Some advocates of DRT have been attracted to the approach because they see it as one that can tell us something about how language is processed by humans. (I was one of those advocates, and I still am up to a point, albeit more guardedly.) But others who have been sympathetic to DRT accounts of linguistic phenomena like nominal and temporal anaphora and who have adopted versions of the theory in their own work – see e.g. (Partee 1984)) or (Geurts 1999)) – have been skeptical of a mentalistic take on DRT. As a consequence, the status of DRT within the wider context of natural language semantics and cognitive science has remained somewhat ambivalent. No such ambivalence is possible for MSDRT, to which we turn in the next section. As indicated by its name, MDRT is explicitly about the mental states of human agents.

Whether or not the form of DRSs has psychological import, however, DRSs are semantic *representations*; they are not just convenient, formal terms that the theory uses to describe semantic values of natural language expressions in models, in the way that Lambda Calculus terms are used in Montague Grammar. As shown by the examples we have discussed, the form of context DRSs plays its part in DRT's account of pronominal anaphora. That DRS form is essential here is demonstrated by Partee's marble example:

(7) a. One of the marbles is not in the bag. It is under the sofa.

b. Nine of the marbles are in the bag. It is under the sofa.

The first sentence of (7.a) and the first sentence of (7.b) have the same propositional content – in every model they determine the same truth values. But nevertheless the pronoun *it* in the second sentence of (7.a) is a perfectly good way to refer to the missing marble, while it is marked or impossible in (7.b).[9]

[9]It is clear that this difference must have to do with the difference in form of the first

DRT is thus a genuine *logical form framework* for doing natural language semantics: It assigns formulas of some formally defined representation language (one of the possible DRS languages) to expressions from the chosen natural language fragment, and the role played by those representations goes beyond that of denoting semantic values. This general architectural feature of DRT is inherited by MSDRT, but with the difference of course that MSDRT's representations are *mental* representations.[10]

3 MSDRT

The original motivation of MSDRT was the conviction that a more fine-grained semantics is needed for attitude attributions than the possible world-based accounts that have dominated formal semantics since the sixties. All these world-based accounts are variants of the proposal made by Hintikka for the semantics of belief and knowledge reports in (Hintikka 1962). Hintikka's proposal is appealing for its remarkable formal simplicity: The belief state, say, of agent A (at some given time t) is characterized by the set Bel(A,t)

sentences of (7.a) and (7.b), for that is the only difference between (7.a) and (7.b). In DRT this difference is captured by the DRSs for the two first sentences. The DRS for the first sentence of (7.a) has a dref for the missing marble in its Universe, which is accessible to the pronoun *it* in the second sentence. The DRS for the first sentence of (7.b) does not have such a dref. (This follows from details about DRS construction that haven't been discussed here, including a precisely formulated rule for the interpretation of anaphoric pronouns. For details see e.g. (Kamp & Reyle 1993).

[10]For more information about DRT: An excellent introduction, by Beaver, Geurts and Maier, can be found in the Stanford Encyclopedia of Philosophy. A detailed introduction to DRT from the early 90s is (Kamp & Reyle 1993), which presents DRT treatments of a fairly broad range of linguistic phenomena, including plurals and tense and aspect. As is plain from this book, most of the hard work in DRT goes into the crafting of DRS construction rules for grammatical constructions and vocabulary items from the natural language fragments to which the framework is being applied. In this regard DRT is like other formal semantics frameworks.
An awkward feature of (Kamp & Reyle 1993) – a complaint that has been voiced repeatedly from the time when the book appeared and one that I would now endorse – is the 'top down' method it adopts for DRS construction, in which sentence DRSs are constructed by starting at the top of the parse tree for a sentence and then stepwise breaking down this tree all the way to its leaves, constructing the DRS in the course of it. (There was a reason for wanting to construct DRSs top down that had to do with the treatment of anaphoric pronouns, but that reason disappeared when anaphora came to be treated as a species of presupposition in the wake of Van Der Sandt's proposal in (Van Der Sandt 1992).) (Kamp et al. 2011) provides an updated survey of DRT which covers the treatment of presuppositions as well as a number of other developments from the time after the appearance of (Kamp & Reyle 1993), including a discussion of MSRT (if not under this name). But there is little in this handbook article about the details of DRS construction.

of all those possible worlds that are compatible with the totality of beliefs that A holds at t. And a belief report of the form 'A believes that ϕ' is true if the proposition expressed by ϕ is entailed by the totality of A's beliefs at t, formally: if the set $[[\phi]]$ of worlds in which ϕ is true is a superset of Bel(A,t). Knowledge attributions are handled likewise, on the basis of the set Know(A,t), consisting of all worlds that are compatible with everything that A knows at t. Later proposals in this general spirit have been more complex. (An example is the proposal that attributions of the form 'A wants it to be the case that ϕ' are true iff worlds in which ϕ holds are (other things being equal) preferred to worlds in which ϕ does not hold (Heim 1992).) But a problem with all such proposals is that they strip semantic content down to something that abstracts from all aspects of the form of the attribution. That is, to put the matter more precisely, they abstract away from the form of the syntactic complement to the attitudinal predicate (such as *believes* or *wants*), which describes what it is that is attributed as belief or desire. And that leads to problems when this form is of the essence to the message that the attribution is meant to convey, as it often is. For instance, many of the attitude attributions we make are made as part of an effort to explain why the attributee did the thing or does the things they did or are doing. But agents do the things for which attitude attributions offer explanations on the basis of bits of practical reasoning from their beliefs and desires and in those bits of reasoning they have to rely on the ways in which they have represented the contents of those beliefs and desires. And attributors try to remain faithful to what they take those content representations to be in how they phrase their descriptions of the contents of the beliefs, desires etc they attribute in their attributions. The forms of those descriptions matter and a credible semantics for such attitude attributions needs to be sensitive to the forms in which attitude contents are described. The possible worlds approach is incapable of that.

MSDRT was developed as an alternative approach that can be sensitive to the form of attitude attributions. One of its ingredients is the Logical Form architecture it inherits from DRT. But it is only one. The Logical Forms that MSDRT proposes for attitude attributions consist of two levels, a first level which offers a description of the relevant parts of the attributee's mental state and a second level at which this description is embedded as one of the arguments of attitude attributing predicate *Att*. The predications involving *Att* are DRS Conditions that can occur in the Condition Sets of DRSs together with DRS Conditions like those we encountered in Section 2. For our use in this paper it is first and foremost the first of these two stages that matters, but certain aspects of the second stage are relevant as well.

3.1 MSDRT, Stage I

In the introduction above attention was drawn to attitude attributions that aim to make sense of people's behavior. Such attributions typically consist of combinations of belief and desire attributions, in the spirit of belief-desire accounts[11] of human action.[12] They describe the mental states of their attributees as involving attitudes of different 'modes' – beliefs, desires and intentions, but also doubts, expectations, queries and more. The mental state descriptions of MSDRT capture this in being sets whose elements are descriptions of attitudes of these various modes. More specifically, we assume that each such element is a pair consisting of (i) a *Mode Indicator* – BEL for 'belief', DES for 'desire', INT for 'intention' and so on (the choice will depend on the way the framework is being applied) – and (ii) a DRS, as specification of the propositional content of the attitude.

But propositional attitude constituents aren't the only constituents of MSDRT's mental state descriptions. A crucial further ingredient of its mental state descriptions are *Entity Representations*. Entity Representations, or 'ERs', as I will refer to them from now on, are representations of things – 'individuals', as the term is used in Model Theory – rather than propositional contents. The motivation for including them in the mental state descriptions that MSDRT provides is two-fold. On the one hand they prove to be needed in dealing with the semantic contributions made by names and other definite noun phrase occurrences within the scope of attitudinal predicates. More often than not the force of such NPs is that the contents of the attributed attitudes are *singular propositions*, which are 'directly' about the referents of those NPs. (More on singular propositions below.) On the other hand there are also more general, less specifically linguistic reasons for wanting to include ERs in our mental state descriptions. There is a strong pre-theoretic intuition that all of us go through life equipped with large inventories of representations of all manner of things – people, places, animals, trees, buildings, works of art, utensils, events of various sorts and even our own thoughts; and entities of all those kinds can either be currently existing or be things of the past. Most of our thoughts are about things for which we have ERs. And since those thoughts are about entities for which we have those representations, they too have singular content.

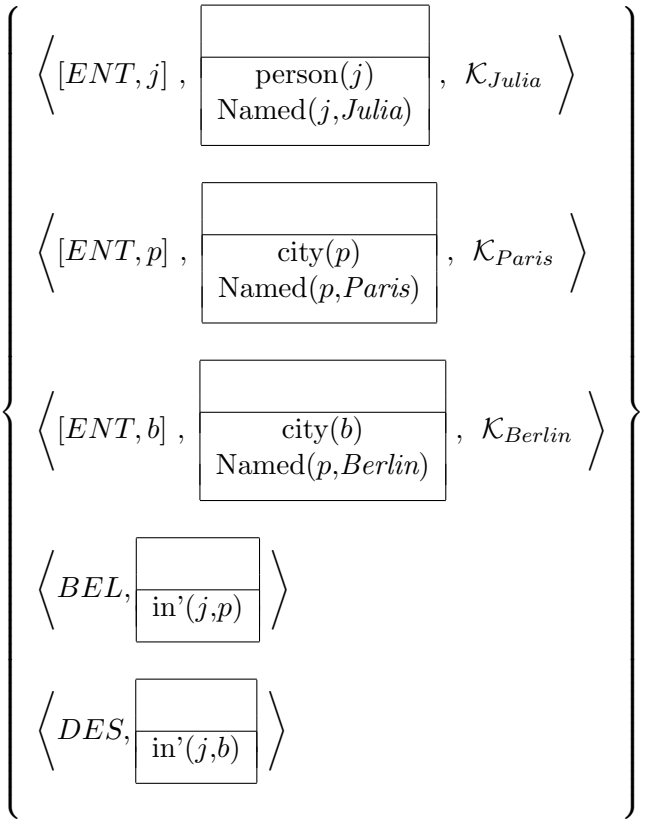
[11] For one reference from a large and complex field, see (Davidson 1963).

[12] In fact, many of the attitude attributions we make extend over several sentences, and often there are anaphoric and other presuppositional relations between the sentences of such attributions. This is a further reason for wanting an account of the semantics of such reports that builds on DRT.

The reason that our ERs afford us singular thoughts about the entities they represent is that ERs are connected to those entities by causal links. ERs 'directly refer' to the entities they represent, to borrow a term from the philosophy of logic and language, because they are linked to them in this manner. That makes it possible for thoughts that involve ERs to be 'directly' about the entities they represent. ERs can thus be regarded as the mental counterparts of directly referring terms in language. I will therefore also speak of ERs as *referring* to the entities they represents and to those entities as their *referents*.

But what is it for a thought to 'involve' an ER? The answer given by MSDRT is that ERs can play the part of arguments in representations of propositional content. More specifically, since propositional content representations are assumed to take the form of DRSs, ERs can act as arguments in DRS Conditions. The way MSDRT has chosen to implement this, staying as close to DRT as possible in this respect, is to assume that each ER comes with a *distinguished discourse referent*. Occurring as an argument in a DRS predication then takes the form of this distinguished dref filling the relevant argument slot of the given DRS predicate.

To see more concretely what MSDRT's mental state descriptions are like, consider the example in (8). This is the description of an agent A who believes that his friend Julia is in Paris and who would like her to be in Berlin (where he is himself). The description assumes that A has ERs for Julia, Paris and Berlin and has the specified belief and desire about the entities represented by these ERs.

(8)


No explanation is needed for the two propositional attitude constituents of (8). Their content representations are DRSs whose form is familiar from what we have seen in Section 2. But note that the Universes of these DRSs are empty. That is because the drefs occurring in their Conditions are imported from the ERs of which they are the distinguished drefs. With regard to the ERs of (8), for now just this. (For more details see Section 3.3.) As can be seen from the ERs in (8), ERs have a good deal of structure, sub-divided into three components. At this point just a couple of hints about these components: (i) The first component consists of (a) the Mode Indicator *ENT*, which tells us that we are dealing with an entity representation (and not some type of propositional attitude constituent) and (b) the distinguished discourse referent of the ER. (ii) The second component contains descriptive information about the referent, which is essential to the way in which the agent conceives of the referent. (iii) The third component is the *anchor set* of the ER. The anchors that are the elements of these anchor sets function as records of causal relations that connect the ER and the agent whose ER it is with the entity represented by the ER. Anchors will play a central role in the story this paper has to tell about causal chains, and will require careful discussion. But that will have to wait till Section 3.3.

I will refer to the mental state descriptions of MSDRT as 'MSDs'.

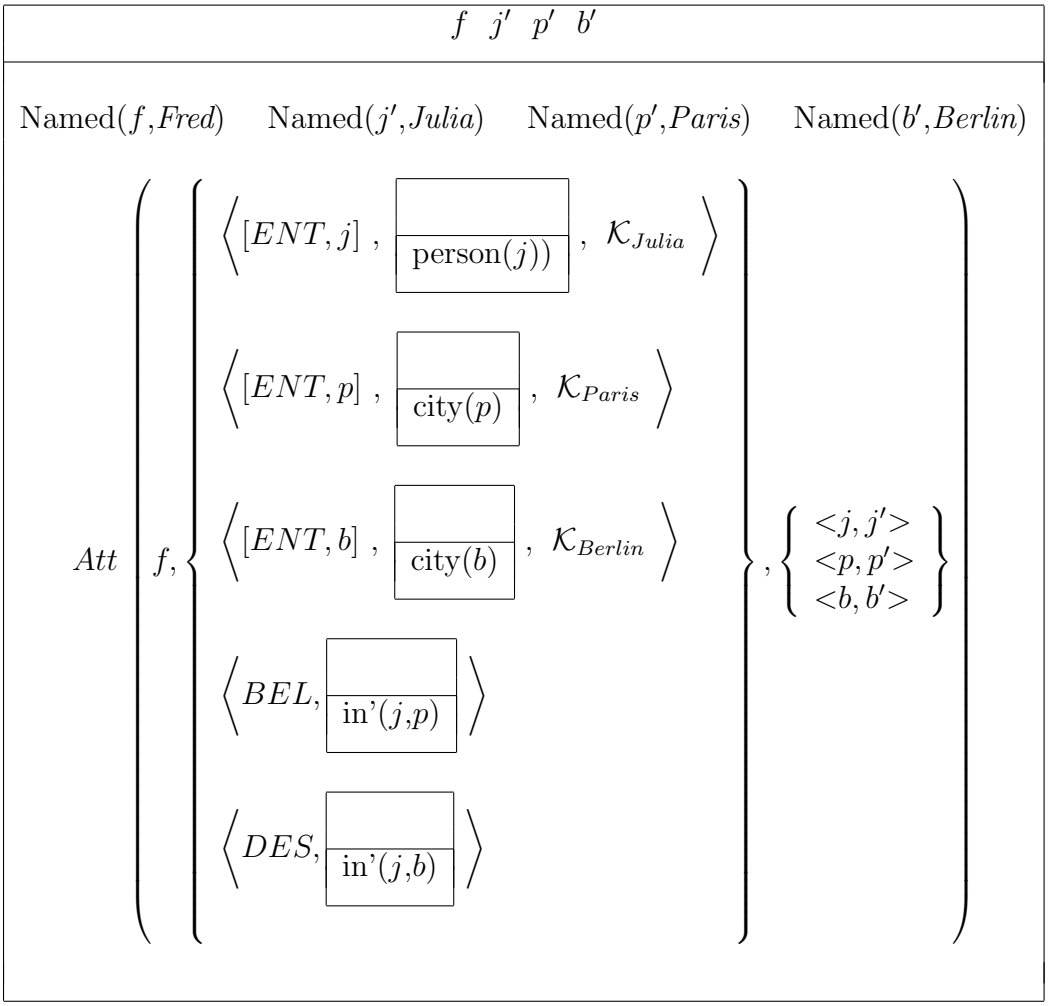
3.2 MSDRT, Stage II

The second stage of MSDRT deals with the Logical Forms for attitude attributions. In these Logical Forms MSDs play a decisive part, but they only occur in special positions, as argument terms of the attitude attributing predicate *Att*. Formally, *Att* predications are DRS Conditions. That makes it possible to combine attitude reports with other kinds of information within one and the same DRS. It also makes MSDRT into a proper extension of DRT. More precisely: like DRT, MSDRT can be thought of as making use of a family of different representation languages and for each of these MSDRT language there is a DRT language of the kind described in Section 2 of which the MSDRT language is a proper extension. An example of such a DRS is the one in (10)[13], the Logical Form for the attitude attribution (9), made by a speaker S who is addressing some person H.

(9) Fred believes that Julia is in Paris. He would like her to be in Berlin.

[13]Throughout this section I ignore, purely for reasons of expository convenience, the times at which agents are in their respective mental states. For instance, in (10) nothing is said about the time at which Fred is supposed to be in the mental state that (9) attributes to him. In the fuller versions of MSDRT that are used in the references given in this section the *Att* predicate has an additional argument slot for a state of affairs s, which consists in the attributee being in the mental state that the *Att*-Condition ascribes to him. The DRS to which the *Att*-Condition belongs can then locate s in time, for instance as simultaneous with the represented utterance of (9).

(10)



Some comments on how (10) is to be read: (1) The predicate *Att* in (10) has three argument slots, (i) for the attributee, (ii) for an MSD and (iii) for a set of *links* which provide referents for ERs from the MSD in the second slot. (2) The links in this set are pairs of drefs $<j,j'>$ and so forth, where the first dref j is the distinguished dref of an ER in the MSD and the second dref j' is external to the *Att* predication. The meaning of the link is that the value of j' (as it gets determined in the course of a model-theoretic evaluation of (10) is the referent of the ER. So the propositional content DRSs that contain occurrences of j – the Belief DRS and the Desire DRS in (10) – evaluate to singular propositions about this value. Note what this entails for the case when (10) is true. (10) entails that Fred, the individual represented by the dref f, is in a mental state that contains an ER with distinguished dref j that is anchored to some individual $\mathbf{j}$. A verifying embedding g of (10) in a model M will have to assign to j' that individual $\mathbf{j}$ from U_M to which this

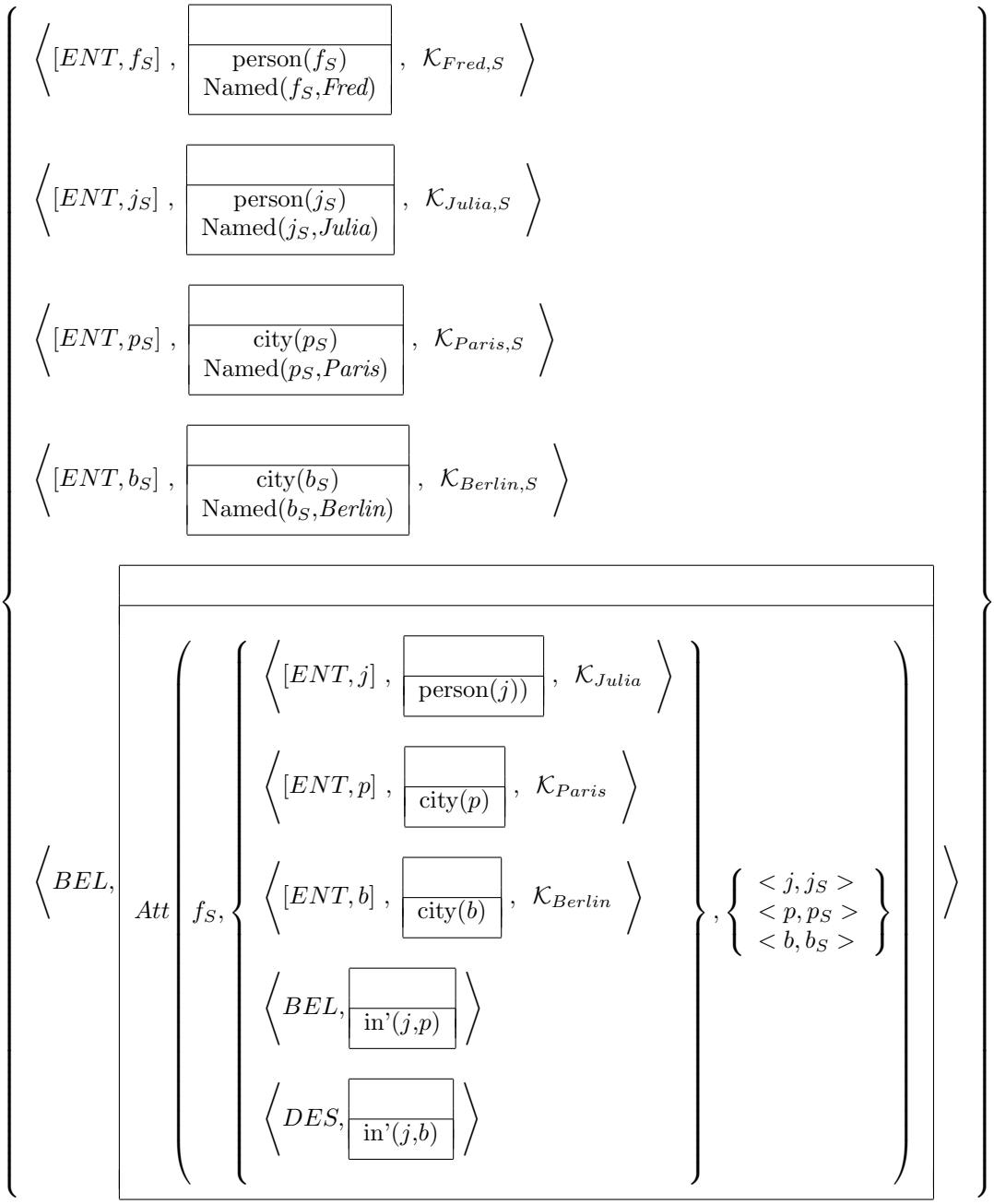
ER is anchored.

The claim made in the last couple of sentences follows from the model-theoretic semantics for MSDRT. That semantics is fairly complex and there is no justification for presenting it here, as the second stage of MSDRT will play no direct role in what follows. (For formal details see (Kamp 2003) and the forthcoming (Kamp n.d.*b*), and for informal discussion of DRSs like (10) see (Kamp & Bende-Farkas 2019).)

DRSs like (10) only give a glimpse of the expressive power of MSDRT. Part of that power derives from the fact that the outer drefs from anchoring links like $<j, j'>$ in (10) can be bound in a variety of ways, including by universal quantification. I will not go into further details here, but readers who have some antecedent experience with DRT may want to play around a little with possible options.

Another source of the expressive power of MSDRT is the possibility of embedding *Att*-Conditions within other such Conditions. So far we have discussed (10) as the Logical form of the utterance in (9) and its *Att*-Condition as the semantic representation of the complex of ERs and attitudes that (9) attributes to Fred. But it is also possible to use *Att*-Conditions as constituents of content DRSs of attitude constituents of MDSs. For instance, if we make the reasonable assumption that the speaker S of (9) believes what (9) expresses, then the relevant part of her mental state could presumably be described by the MDS in (11), consisting of ERs for Fred, Julia, Paris and Berlin togehter with a belief whose content representation consists just of the *Att*-Condition in (10):

(11)



In (11) the anchoring links are between the distinguished drefs of the ERs in the content DRS of S's belief and the distinguished drefs of S's own ERs. This guarantees that the propositional contents of the attitudes ascribed to Fred by the MSD that occurs as argument to the *Att*-Condition denote the propositions that S wants to attribute to Fred: singular propositions about the referents of her ERs for Julia, Paris and Berlin. A similar point can be

made about what presumably happens to the mental state of H as a result of his interpretation of S's words in (9). Let us assume that H too has ERs for the Fred, Julia, Paris and Berlin that S uses in (9) and that these ERs are labeled with the names *Fred*, *Julia*, *Paris* and *Berlin*, respectively, and that he uses these ERs in his interpretation of (9). If he furthermore constructs his interpretation according to the rules of the language, obtaining a representing DRS for (9) that is in essence like (10), and if he also believes what this DRS tells him, then the relevant part of his resulting mental state may be assumed to look just like (11). The contents represented by the attributed belief and desire in this part will now be tied to the referents of *his* ERs for Julia, Paris and Berlin. But if these ERs are properly coordinated with the corresponding ERs of S, in the sense that they refer to the same referents, then the belief and desire attributions to Fred by H and S will have the same propositional content; and with that the propositional contents that S and H associate with (9) as a whole will be the same as well.[14]

This concludes the high speed introduction to MSDRT, except for the discussion of the internal structure of ERs in the next two subsections. To repeat once more what I have emphasized at various points: The formal foundations of the framework – DRS construction rules, syntax and model-theoretic semantics for MSDRT's DRS-languages – have been passed over, but can be recovered from the documents mentioned along the way.

3.3 Entity Representations and their anchors

Some things about the structure of ERs have been noted in passing. But what we need is a formal definition of MSDRT's concept of Entity Representations.[15]

[14]Often agents will have several ERs labeled by the same name. For instance, both S and H may have more than one *Julia*-labeled ER: they each know more than one person with the name *Julia*. If that is so, H should use the *right Julia*-labeled ER when he interprets (9) – that ER which is coreferential with the *Julia*-labeled ER that S made use of when choosing the name *Julia* as part of her utterance of (9). Interpretational mishaps, when the interpreter uses a wrong N-labeled ER to interpret the speaker's use of N, are in principle always possible, and happen not infrequently. It is of course also possible that although H has one or more *Julia*-labeled ERs, none of those is coreferential with the ER that S has used. For discussion of various such cases see (Kamp 2015) and also footnote 24 in Section 3.5. and Section 4.

[15]There is an extensive literature on entity representations. In much of it such representations go by the names of 'file card' (Perry 1980), (Heim 1982,1988) or 'mental file' (Recanati 2012). I expect that many readers will be familiar with a good part of this literature and that those who have been wondering to what extent the notion of an *Entity Representation* as it is used in MSDRT is a misappropriation, barely concealed by a

(12) **Definition of *Entity Representation***

An *Entity Representation* (*ER*) is a triple $<[ENT, \alpha], K_{descr}, \mathcal{K}_{anch}>$, where

(i.a) ENT is a Mode Indicator, indicating that the mental state constituent to which it belongs is an Entity Representation, as distinct from the various Modes of the possible propositional attitude constituents of MSDs;

(i.b) α is a discourse referent (the *distinguished dref* of the ER).

(ii) K_{descr} is a DRS (which contains certain descriptive information about the represented entity).

(iii) $\mathcal{K}_{anch}$ is a set of *anchor-DRSs*[16]

No further comments are needed for the first ER-component. As regards the descriptive information in the second component just this: The guiding intuition is that information in this component is restricted to what the agent treats as non-contingent information about the referent, as she conceives of that referent by virtue of her ER for it. It is not easy to be more precise on this point, for one thing because agents often waver in their conceptions of entities – what the agent treats as non-contingent information about the

new (rather uninspired) label. There is some justification for this suspicion. The use of files and the file cards in the philosophy of language has been known to me from before the first formulation of DRT, first through the work of Perry and then through the File Change Semantics that Heim developed in her doctoral dissertation. At the time when File Change semantics and DRT were made accessible to a wider public, there was a discussion about the relations between Heim's file cards and DRT's discourse referents. Of the various notions of entity representations that can be found in the literature and the roles they play in the different theories that make use of them the one to which the ERs of MSDRT bear the closest resemblance are the *mental files* defined and used in the work of Recanati. (See in particular (Recanati 2012), which discusses many examples and puzzles from the philosophy of language and mind literature to which his concept of a mental file can be fruitfully applied.)

The reason for not sticking with the file card/mental file terminology and adopting the label 'Entity Representation' was to avoid unwanted identifications with notions from the file literature that are not developed in the formally explicit way that ERs are treated in MSDRT. (The formal explicitness of MSDRT is also found in Maier's *Attitude Description Theory*, see e.g. (Maier 2016).)

[16] For those readers who are familiar with some of the existing publications in which MSDRT is used: The anchors that occur in anchor sets are often referred to as *internal anchors* in those publications, while the term *external anchor* is used to refer to what in the present paper is described as 'the entity represented by the ER' or as the ER's 'referent'. In the present paper the terms *internal* and *external* will only be used in the discussion of anchors in this section.

referent at one time or in one context need not be the same as what she treats as non-contingent information about the referent at other times or in other contexts.[17] In this paper I make no assumptions about what information goes into the second components of ERs except that this information includes (i) 'sortal' information, which specifies what sort of entity the ER's referent is (a person or a city or an artifact and so on), and (ii) information about the name or names that the agent assumes the referent goes by. What else may or may not go into the second components of ERs is of no consequence for the remainder if this paper.

What does matter crucially for present purposes are the third components of ERs, their anchor sets. The remainder of this section is about them.

For a start, it is important to keep firmly in mind that the anchors that make up the anchor sets of ERs are *internal* anchors in the sense of footnote 16; they are *records* of causal relations between the ER and its referent.

Second, the anchors that can be found in anchor sets come in different types, reflecting the different causal relations to which they bear witness. One of those types are the *vicarious anchors*. These are the anchor types that are of principal interest in this paper, as the pivots of our reconstruction of causal chains. However, the idea behind the notion of anchors as constituents of Entity Representations is more easily explained by first looking at another type, that of the *perceptual anchors*.

One situation in which an agent can form a new ER is when she perceives something that she takes to be something that she doesn't know. Such an ER comes equipped with an anchor set containing a single anchor, which records the perception that results in the formation of the ER. Such anchors, which act as records of perceptions of the entities represented by such ERs, are called *perceptual anchors*.[18] But they are not the only ones. Perceptual

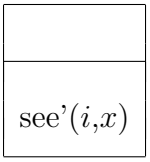
[17]Contingent information about the referent of an ER may be assumed to take the form of belief constituents of the agent's mental state, whose content specifying DRSs have occurrences of the distinguished dref of the ER. For some discussion of MSDRT's distinction between treating information as contingent and treating it as non-contingent see(Kamp 2021), Sn. 4.2.
The idea that what we take as essential to entities of which we have a conception can vary with the context in which we think about the entity seems to have been an important ingredient in the so-called 'cluster theory of names'. See *Naming and Necessity*, (Cumming 2019).

[18]Sometimes our perceptions are non-veridical. We can be the victims of fata morganas can other kinds of perceptual illusions. Since agents cannot distinguish such 'fake' percep-

events can also be the records of *perceptual recognition events*, in which the agent perceives an entity that she takes to be the referent of an ER she already has. This is one of the reasons why MSDRT assumes that ERs have anchor sets, which can consist of several anchors rather than just a single one.[19]

In applications of MSDRT it is often important to define anchor sets explicitly, and for that we need explicit descriptions of the anchors they contain. To this end MSDRT has been making use of DRSs from some specialized anchor representation language. (13) is a simple example of such a DRS.

(13)



It identifies the anchor of an agent who is currently looking at an entity, represented by the distinguished dref x of the ER to whose anchor set (13) belongs. (Non-visual perceptions will of course require other perception predicates than *see'*.) After the perception is over, but the ER continues to be part of the agent's mental state, (13) will be adjusted to an anchor form which expresses that a visual perception of the entity *has* occurred in the past, possibly with some specification of when that was, but we do not need

tions from true ones, it may be assumed that the fake perceptions lead to the formation of ERs with perceptual anchors just as real perceptions do. But in such cases the perceptual anchor of the ER is a false witness, the ER has no referent and attitude constituents of the agent's mental state whose content representations have occurrences of the ER's distinguished dref fail to have a well-defined propositional content. Important as this aspect of MSDRT may be, I have decided to set it aside in this paper, to reduce expositional overload.

[19] A consequence of this way of treating recognition in MSDRT is that the anchor sets of ERs for entities that one interacts with on a regular basis and that one thus encounters again and again would grow to very large proportions. From a psychological point of view this seems unrealistic. The problem will be addressed in a forthcoming paper (Kamp n.d.*a*).

Another problem connected with the use of ERs in recognition is the information that must be associated with the ER in order that it can play its part in the recognition process. For one much discussed type of case consider facial recognition of people. This is something that most of us are remarkably good at. (And that algorithms are now getting very good at too, though I do not know if that has any relevance for what I want to say.) In order that my ER for you can do its job when I see you or a picture of you and see that it is you, my ER must have some kind of mental image associated with it (one that I will also be able to call up before my mind when you are not there). What precisely is associated with the ER that makes such recognitions possible and how it is associated with the ER are questions about which MSDRT has had nothing to say so far. See footnote 30.

to go into the details of that here.

So much for perceptual anchors. More important for the purposes of this paper are as I said vicarious anchors, the topic of the next section.

3.4 Vicarious anchoring and vicarious anchors

One of MSDRT's central claims is that ERs can be linked to their referents via *vicarious anchors*. Agents can achieve direct referential access to entities just on the strength of the linguistic references made to them by others – this was one of the lessons I first learned from *Naming and Necessity*. But the implementation of the message in MSDRT looks somewhat different from the way I understood it at the time, and it also is somewhat more general. I should add that the message – that direct access to entities can be obtained from the references made to them by others – is not one that can be inferred from any other assumptions made in MSDRT. But while it could be separated from the rest of it, MSDRT would be a very different enterprise without it.

As a first example of vicarious anchoring and vicarious anchors consider the mini-conversation in (14).

(14) S: I just reread a short story by Gogol last night. Amazingly good, really!

H: Ah yes, interesting!

This exchange is about as innocuous as they come. But it could plausibly occur in a number of different situations, and it is the difference between some of those situations on which we focus. One difference has to do with H. The name Gogol will either be familiar to H when the conversation starts or it won't be. In the first case a 'Gogol'-labeled ER for Gogol will be a constituent of H's mental state, in the second case it won't. A second question has to do with what S thinks about H. S herself will of course have a 'Gogol'-labeled ER for Gogol, otherwise she could not form the thought expressed by her words and express that thought the way she does in (14). But she may also have an idea of whether H has such an ER as well; and if she is using the name 'Gogol' in the way she uses it in (14) she better should. For 'standard' uses of proper names like this one carry a presupposition that the recipient is familiar with the name – in our terms: that he has an ER labeled with the name for the entity that the speaker is referring to by using the name. Only when H has such an ER will he be in a position to interpret A's use of *Gogol*

in the way such name uses are meant to be interpreted. If H doesn't have a *Gogol*-labeled ER for Gogol, then he must accommodate the presupposition of S's use of *Gogol* by adopting one afresh, as if he had had one all along.[20]

But if H doesn't have a *Gogol*-labeled ER for Gogol and needs to accommodate one, what should that ER be like if it is to play the part in his interpretation of (14) that would have been played by a *Gogol*-labeled ER for Gogol if the name had been familiar to H and he had had such an ER from the start? The answer to this question is one of *Naming and Necessity*'s important messages: H can take over S's use of *Gogol*, as name for the individual that S has just used *Gogol* to refer to, from the use that S has just made of it in the utterance that reaches him. In the terms of MSDRT: H can mark the ER he forms as one that refers to whatever it is that S referred to with her use of *Gogol*. Such marks are called *vicarious anchors*.[21]
By using the distinguished dref of his new vicariously anchored ER for Gogol H is in a position to form a representation of S's statement S that correctly captures its propositional content – a representation that expresses the proposition about Gogol that S read one of his short stories the night before. That is secured by the vicarious anchor of his ER, which renders that ER coreferential with the one relied on by S. But the vicarious anchor of H's ER also accomplishes something else: it establishes a *Gogol*-related *correlation* between H and S. As we noted, it is one of the assumptions of MSDRT that to use a name N in the manner that S uses *Gogol* in (14) you must have an N-labeled ER that represents the referent that you are using N to refer

[20]Sometimes people will use names (in the 'standard way', see (Kamp 2015)), when they don't believe the name isn't familiar to their addressee, but want to impress or intimidate them. Or they really have no idea whether their addressees are familiar with the names they are using, but simply don't care. Related to cases of this last sort are those where the names that interpreters are confronted with are used in conversations they overhear but which are not intended for them. See (Evans & Altham 1973).

[21]Whether just one such exposure suffices to give H a proper command of *Gogol* has been a point of long-standing debate. For an early discussion of this point see (Evans & Altham 1973). To relate the point to (14), suppose that H has never heard of Gogol.
Perhaps the most natural situations for some of us to acquire what feels like a genuine command of a name are our encounters with unfamiliar names in texts. Suppose I read in a textbook: "The first proof that combustion involved the binding of oxygen was given by Lavoisier." I never heard of Lavoisier before and so one of the tings I learn from this sentence is that there was a person by this name who had something to do with chemistry. (If I am conscientious student, I will make a note of the name as well as the information that the passage provides about oxygen.) In this case too I may feel I need to learn more before I feel confident in using the name myself. But it seems quite plausible that just on the strength of my first encounter with the name *Lavoisier* I set up a labeled ER for its referent.

to. So if S made a proper use of *Gogol* in (14), then she had a *Gogol*-labeled ER ER_S for the Gogol that she was referring to on this occasion. And when H interpreted S's use of *Gogol* by accommodating an ER ER_H for the individual referred to by her, and that is represented by the *Gogol*-labeled ER ER_S that her use of *Gogol* was based on, then the vicarious anchor of ER_H can be seen as creating a link between their ERs ER_S and ER_H. We will represent the link as the ordered pair $<ER_S,ER_H>$ and see it as standing for a coordination relation between S and H, to the effect that their ERs ER_S and ER_H are represent the same individual.

When the recipient already has an N-labeled ER ER_H for the referent that the speaker is referring to by means of N and uses it to interpret the speaker's use of N, then no vicarious anchor is needed to establish coreference between ER_H and the ER ER_S on which her use was based. But the use H makes of ER_H in his interpretation of N has a coordinating effect in this case too. It won't normally establish the coordination between their ERs, but it will often strengthen the coordination, for instance by making it more direct. To do justice to these coordination effects, MSDRT assumes that in such cases too the use that H makes of his ER ER_H to interpret S's use of N, involves the adding of a vicarious anchor to the anchor set of ER_H, thereby also creating a link between ER_H and the ER ER_S on which the speaker's use of N was based. So in these cases a link $<ER_S,ER_H>$ is created as well, as a signal for the reinforcement of the coordination relation due to the given use and interpretation of N, between S and H and more generally within the N-using community.

To summarize the main points of this section:

(15) (i) When a speaker uses a name in the 'standard' way exemplified by the use of N by S in (14) then she must have an N-labeled ER on the basis of which she does this, and by basing her use of N on this ER the entity she refers to by her use is the one represented by this ER.

(ii) A member of the speaker's audience who interprets her use of N must either use an N-labeled ER that he has or else accommodate an N-labeled ER as representation for the entity that the speaker has referred to with her use of N. In either case a vicarious anchor is added to the anchor set of the ER as a witness of the ER's use in the interpretation of N on the given occasion of its use.

The vicarious anchor enforces coreference between the interpreter's ER

ER_H and the speaker's ER ER_S. Furthermore it adds a new link $<ER_S,ER_H>$ to the coordination between S and H withv regard to their use of N.

The links $<ER_S,ER_H>$ will be the pivots in our reconstruction of the causal chain theory. This is the topics of Section 4. In Section 3.5, the last part of the present section, we look at a couple of examples of how utterance interpretation affects the mental state of the interpreter.[22]

3.5 The effect of utterance interpretation on mental states

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

[22]The term 'vicarious' is familiar from the literature on file cards and mental files, though its meaning varies between the different theories that make use of it. A prominent use of 'vicarious' is made in (Recanati 2012). It is one with which the use of 'vicarious' in MSDRT has much in common, although the two uses differ formally in that they apply to different types of entities: Recanati speaks of 'vicarious mental files', whereas in MSDRT 'vicarious' is a predicate of ER anchors. But there is nevertheless a close connection between the two. Recanati's vicarious files, to which he also refers as 'indexed files', are files that one agent A_1 may have for a file in the mind of another agent A_2. Formally, a vicarious/indexed file is a pair $<a_2,f>$, where (i) a_2 is a 'term in the language of thought' of A_1 that refers to A_2, and (ii) f is a term of that language that refers to a mental file of A_2. In order that f can refer to a file in the mind of A_2 there must of course be such a file. (What further properties the term f must have if it is to refer to a file of A_2 with the relevant properties is a matter for which the reader should consult (Recanati 2012). For the relation between vicarious files and vicarious anchors the details are not essential.) Furthermore, Recanati assumes that if A_1 has an indexed file for a file of A_2, he will also have a regular file (i.e. a non-indexed file) for the referent of this file of A_2. And these two files, A_1's regular file and his indexed file, are said to be 'vicariously linked'.
In the version of MSDRT presented here there is no direct counterpart to Recanati's indexed files. But for the cases that are relevant to this paper – those where A_1 interprets an utterance of a name N by A_2 – the role played by the vicariously linked pair of regular and indexed files in the mind of A_1 and the file in the mind of A_2 that A_1's indexed file refers to is played by the A_1's vicarious anchor. The formation of this anchor presupposes that A_2 had an N-labeled ER for the referent of N on which her use of N was based, and by adding this anchor to his N-labeled ER A_1 confirms this presupposition, and formally establishes a link between his ER and hers. (For more details see Sections 3.5 and 4.) No provision is made in this account for a representation in A_1's mind of the ER that A_2 must have used in her production of N. That may feel that this is a shortcoming of the MSDRT account given here. But there is more than one way to fill this gap. One is to stipulate that the mental state of A_1 which results from this interpretation of N should contain the belief that A_2 has an ER on which her use of N was based. (16) is a proposal for the MSD for this part of his mental state.

interpreter. We will be looking at just one sentence, a streamlined version of S's utterance in (15):

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

We consider two scenarios in which S's uses and H interprets (17). In the first H doesn't have an ER for Gogol, in the second he does. (18) gives the MDS for the relevant part of the mental state of S, assuming that S herself believes the information that is expressed in (17) and that she has a *Gogol*-labeled ER for Gogol that puts her in a position to use the name *Gogol* in (17). (19) describes the relevant part of H's mental state in the first of our two scenarios just before his interpretation of (17) gets under way. Note that the only ER shown in (19) is an ER for the speaker S. The fact that (19) does not display an ER for Gogol is to be understood in this case as an indication that the mental state partially described by (19) doesn't have such an ER.

(18)
$$\left\{ \begin{array}{l} \left\langle [ENT, g_S] \ , \ \begin{array}{|c|}\hline \\ \hline \text{person}(g_S) \\ \text{Named}(g_S, \textit{Gogol}) \\ \hline \end{array} \ , \ \mathcal{K}_{\textit{Gogol}} \right\rangle \\ \\ \left\langle \textit{BEL}, \begin{array}{|c|}\hline e \quad y \\ \hline e \prec n \quad \text{short-story'}(y) \quad \text{by'}(y, g_S) \\ e\text{: read'}(i, y) \\ \hline \end{array} \right\rangle \end{array} \right\}^{23}$$

(16)
$$\left\{ \begin{array}{l} \left\langle [ENT, a_2] \ , \ \begin{array}{|c|}\hline \\ \hline \text{person}(a_2) \\ \hline \end{array} \ , \ \mathcal{K}_{A_2} \right\rangle \\ \\ \left\langle [ENT, r] \ , \ \begin{array}{|c|}\hline \\ \hline \\ \hline \end{array} \ , \ \mathcal{K}_r \right\rangle \\ \\ \left\langle \textit{BEL}, \begin{array}{|c|}\hline er \\ \hline \text{ER}(er) \quad \text{Belongs}(er, a_2) \quad \text{Repr}(er, r) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

New to the MSD language used in (16) are (i) the sortal predicate 'ER', which is true of all and only mental state constituents that are Entity Representations, (ii) the 2-place predicate 'Belongs' which holds between agents and constituents of their mental states and (iii) the 2-place predicate 'Repr', which holds between ERs and their referents.

[23] Notation: 1. The symbol 'g_S' that has been chosen here for the distinguished dref

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

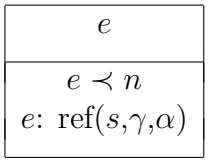
When H is in the mental state described in (19), then his interpretation of (17) requires the accommodation of a vicariously anchored ER for Gogol. The result of this accommodation and the construction of H's representation of the content of (17), which, we assume, H accepts as true as well and thus adopts as a belief, leads to the mental state described in (20).

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

Unfamiliar in (20) is the notation used for the vicarious anchor of H's accommodated ER for Gogol. Like for perceptual anchors, some suitable form is needed to represent such anchors. The form adopted here is shown in (21).

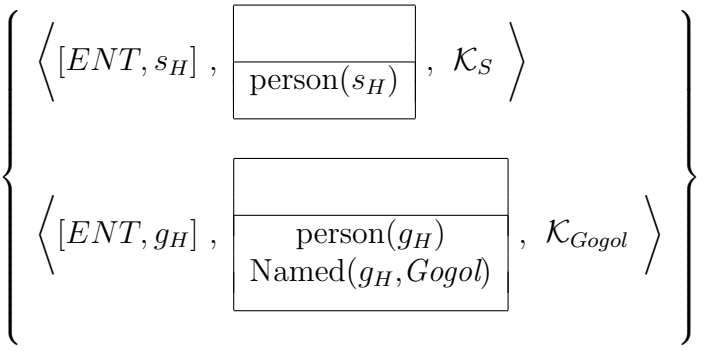
---

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 (21) below. (The subscripts $_S$ and $_H$ are chosen solely for mnemonic reasons – $_S$ means that we are dealing with a dref in an MSD for S, and $_H$ likewise that we are deaing with a dref in an MSD for H. They have no theoretical import.) 2. The content DRS of the belief in (18) belongs to a DRS language that extends beyond the description of DRS languages in Section 2.1. 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) – see also the English version (Kamp 2017) – and in (Kamp & Reyle 1993). The intuitive meaning of the DRS in (18) 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 (18) describes.

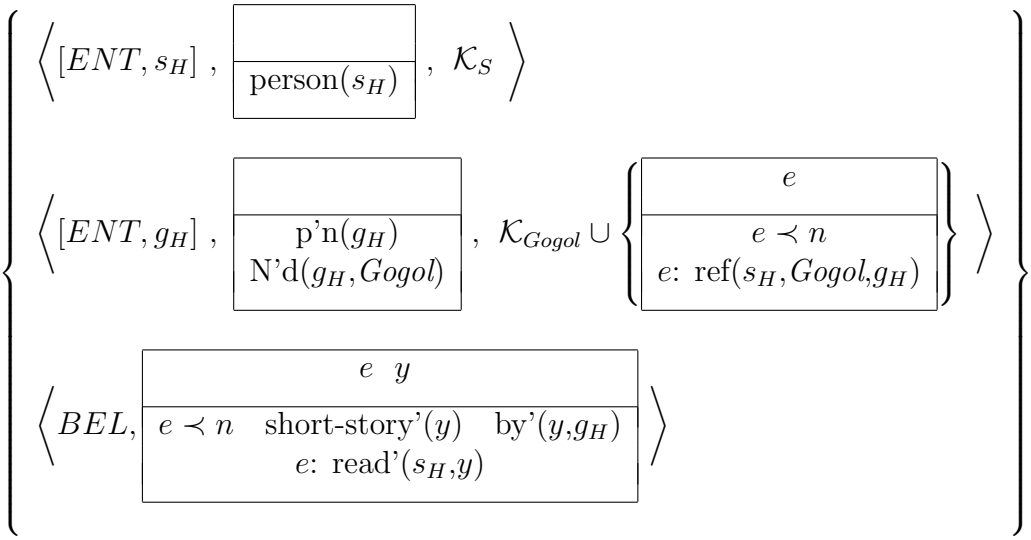
(21)


In (21) 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 (21) is an anchor. (Thus in the vicarious anchor of the ER for Gogol in (20), 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 (17) to him. In this case H's mental state before he starts his interpretation of (17) can be described as in (22).

(22)


According to MSDRT the result of processing (17) is the one shown in (23).

(23)


The one aspect of (23) that requires attention is the anchor set of its *Gogol*-labeled ER. This set results from adding to the anchor or anchors that were part of the ER before interpretation started the same vicarious anchor that figures as the unique anchor of the accommodated ER in (20). The reason I gave in the last section for insisting that in this case a vicarious anchor is introduced into the anchor set too (and not only when a new ER is accommodated) was that also when the interpreter uses a previously adopted ER, the communication reinforces the coordination between him and S. But a further justification could be given as well. When H uses an ER he already has, this means that he is understanding the name that S is uses for someone familiar to him – he knows who Gogol is, the famous author of the "Dead Souls", and understands that S is saying something about that author. In a sense that is a form of recognition too, in certain ways quite like what happens when you recognize someone you meet in the flesh. In both cases the experience strengthens your relationship with the referent. So it is natural that just as we assumed that visual recognition of an entity **r** for which one has an ER adds a perceptual anchor to the ER's anchor set, so you add a vicarious anchor to the anchor set when you establish contact with **r** by hearing or reading their name.

In fact, anchor sets are quite often mixtures of anchors of either type. Take for instance the case of Victor, who sees a woman at a party with whom he is much taken and tries to find out more about her. When someone tells him her name, he will add a vicarious anchor to his thus far purely perceptually anchored ER for her (as well as the Condition 'Named(x,N)', where x is the distinguished dref of the ER and N the name he has just been told). Or conversely, take Pierre, who grows up in Paris and learns about London from the uses that his family makes of the name *Londres*, thus obtaining an *Londres*-labeled ER for London with a (possibly quite large) set vicarious anchors. When he then finally gets to London and recognizes it as the place for which thus far only had an ER whose anchor set was purely vicarious, that will from then on be boosted with a growing number of perceptual anchors.[24]

[24] Of course, it is also possible – even if the possibility is a somewhat outlandish one – that Pierre doesn't recognize London as the referent of his vicariously anchored ER for London when he finally gets there. Instead he will form a new ER for London, with multiple perceptual anchors and, as in the story of (Kripke 1979*a*), also vicarious anchors that connect him with people who have been referring to the place where he is now by using the name *London* in his presence. So now Pierre has two ERs for London, both of which are labeled, but with different names, and he lacks the resources to put one and one together. (What we are to say about belief attributions that can be made to Pierre in which the names *London* and *Londres* are used is a difficult matter, as we have learned from Kripke's paper, one of the harder nuts that theories of the use of names in attitude

It is time to summarize the main points of this and the last section on vicarious anchors and their reference-fixing and coordinating roles. Vicarious anchors have the following two functions:

• Vicarious anchors secure proper reference for the ERs of which they are part, by themselves or in cooperation with other anchors with which they cohabit in their anchor sets. This is so in particular for the vicarious anchors of ERs that are accommodated in response to unfamiliar names and that are the sole members of their anchor sets.

• A vicarious anchor links its ER to the relevant ER or information source of the reference that the anchor records. These links are the constituents of the intersubjective networks that stabilize the use of names between language users and within the language communities to which they belong.

---

attributions should be able to crack. But it is not one of the nuts to be cracked in this paper.)

I should add that cases of recognition failure, where you are confronted with an entity for which you have an ER, but do not realize this, are only one way in which we can be making a suboptimal use of our ERs. Another form of sub-optimality, and in this case we can speak of an outright mistake, is when there is too much recognition: you think that whom you see on the other side of the street is Julia, adding a perceptual anchor to your ER for Julia (as well as, potentially more seriously, associating new descriptive information with that ER that derives from your present encounter). According to the account presented here this renders your ER *incoherent*, in that different anchors in its anchor set connect it to different entities. Trying to refer to two things by virtue of being connected to both is hardly better than failing to refer because of not being connected to anything. Still, mishaps of this kind often remain without serious consequences. When they do and when they cause serious havoc is yet another topic for further investigation. Recall the brief discussion of this issue in footnote 14.

It is tempting to add an MSDRT analysis of the predicaments of Kripke's Pierre at this point; but it is one that I resist as it would lead us away from the main story. From what has been said in this paper so far, however, it is probably not too hard to see how at least part of such an analysis would go: Pierre has a *Londres*-labeled and a *London*-labeled ER for London. The distinguished drefs of those ERs may occur in the content representations of various attitudes of his, without any logical constraints on their mutual distribution. In particular it is possible for Pierre to hold beliefs with the content specifications 'beautiful'(l_f)' and 'ugly'(l_e)', where l_f the distinguished dref of his *Londres*-labeled ER and l_e the distinguished dref of his *London*-labeled ER. The propositional contents determined by these specifications are mutually exclusive since the two ERs are coreferential. But to realize that they are incompatible one has to be aware of their coreferentiality and Pierre isn't aware of that. Everyone who has read (Kripke 1979*a*) knows there is more to be said about this problem, but this is how the first part of a MSDRT-based analysis would go.

• As I read what *Naming and Necessity* has to say about causal chains, MSDRT differs from that in assuming that every interpretation of a reference by a proper name produces a link between interpreter and source, irrespective of whether the name is new or familiar to the interpreter. In the latter case a link is established between language users both of whom already have command of the name.

• Furthermore, vicarious anchoring is not restricted to the interpretation of reference events involving proper names, but can result from the interpretation of other referring noun phrases as well.

Let me give one illustration of this last point. When my mother told me when I was a teenager about the apartment in Charlottenburg where she spent the first nine years of her life, she must have had, by any of the criteria I can think of, a multiply anchored ER for that apartment. And through what she was telling me about the apartment she enabled me to form an ER of that apartment as well, one that I reused again and again, when I listened to more of her stories about it or indulged in my own fantasies about it. But my ER only had vicarious anchors, and it was without a name. (If I had known the address of the apartment, that might have played a role much like that of a name, but I never knew it, and now regret I never asked.) The NPs my mother used to refer to the apartment were descriptions like *our apartment*, descriptions she used to refer to the entity represented by her ER. Such uses of definite descriptions are common; and other noun phrases types, e.g. demonstratives and pronouns, can also be used to refer to entities represented by one's ERs. Vicarious anchors, MSDRT assumes, result from the interpretations of any such referential acts.

4 ER Networks

The main gist of the story will now be clear: Vicarious anchors establish links between ERs of different language users, the ER to which the anchor belongs and the ER on which the speaker relied in producing the reference that the anchor records. It is such links that form the chains which link any speaker who has command of a name N as name for a referent $\mathbf{r}$ to someone who participated in the 'baptismal event' in which N was conferred as name upon $\mathbf{r}$. But there is more to say.

The causal chains that link N-users with baptism participants are normally linear substructures of ER networks that are much bigger, and that also tend

to have a much more complex structure. We start with a look at the links from which such networks are built.[25]

Vicarious anchors, I said towards the end of the last section, arise not only in response to referential uses of names but also when sources refer through the use of other types of definite noun phrases.[26] When the interpretations of referring acts involving these other types of noun phrases give rise to new ERs, these ERs will not be named (i.e. not be N-labeled for any name N). The vicarious anchors of such ERs will nevertheless establish links $<ER_S,ER_H>$ between the new ER ER_H of the interpreter and the ER ER_S of the speaker who made the reference. In such cases ER_S can be either named or unnamed: S may not have a name for the entity she is referring to, but it is also possible that her ER is N-labeled, but that she prefers to refer to the entity it represents by using some other type of noun phrase, for instance because she thinks that this will make it easier for H to understand which entity she is referring to. And the opposite situation, where ER_H is named but ER_S is not, is possible too, for instance when S doesn't have a name for the entity and that she is looking at while making her reference, and where H, who can also see the entity, recognizes it as the one represented by an N-labeled ER he has for it. The most general notion of an intersubjective coreference network is that which consists of all such links $<ER_S,ER_H>$ (Kamp 2021). In this paper, however, we restrict attention to the sub-networks that consist of links $<ER_S,ER_H>$ where ER_S and ER_H are both N-labeled for the same N and that are the result of interpretations by H of references made by S through the use of N.

Before we move on with the discussion of N-labeled networks and causal chains, there is something I want to make fully explicit, although it may have become clear enough from the discussions up to this point. It is about the identity of names. I am assuming that the identity of a name is merely a matter of morphophonemic form. If your friend is called *Julia* and my friend is called *Julia*, then your friend's name and my friend's name are the *same* name, which your friend and my friend share. The name *Julia* is ambiguous between your friend and my friend and the countless other people who are also called *Julia*. When you say to me "I just talked to Julia", I may misun-

[25]For a discussion of ER networks that partly overlaps with the one in this section see (Kamp 2021).

[26](Kamp & Bende-Farkas 2019) argues that this is so even when speakers make epistemically specific uses of indefinites and their interpreters take them to be doing that. For definitions of epistemically specific and other types of specific uses of indefinites see (Farkas 2002).

derstand you by taking you to be talking about my friend, whereas you were talking about your friend. For me your statement is ambiguous: among my *Julia*-labeled ERs there are at least two, the one for my girlfriend and the one for your girlfriend: both are plausible candidates for my interpretation of the use you are making of *Julia*. On this conception of the identity of names, all names are potentially ambiguous, quite a few names *are* ambiguous and some are ambiguous to a very high degree.[27]

Where there is ambiguity there is always a risk of misinterpretation. The misinterpretations caused by the ambiguity of a name N take the form of S using N to refer to the entity represented by one of her N-labeled ERs and H taking her to be referring to some other entity, represented by some N-labeled ER ER_H that he has. When this happens, the damage is two-fold. On the one hand the interpretation renders ER_H incoherent in that its anchor set now contains divergent anchors – the one or ones that link it with the entity for which ER_H was originally introduced and the new vicarious anchor that links it to the entity that S has referred to (see footnote (24)). On the other hand, H's misinterpretation also produces a deficient link $<ER_S,ER_H>$ – there isn't any **r** such that ER_S and ER_H are both unequivocal representations of **r**. Such deficient links are a threat to the networks to which they are added: it ought to be the case that any two ERs occurring in the network that are connected by a chain of links are unequivocally coreferential, but for networks with deficient links this isn't so; and when the ERs from the deficient link are both N-labeled, this may cause upheavals in the subsequent use of N. (For some discussion see (Evans & Altham 1973)).[28]

To conclude this discussion of link deficiency, note that the use of a name N can never lead to a deficient link when the interpreter doesn't have an N-labeled ER to interpret the speaker's use of N. That is, in those cases where referring by means of N has the effect of introducing the addressee to the given use of N there is no room for misinterpretation and therefore also no risk of it. The causal chains that, on my understanding, *Naming and Necessity* is speaking of – those consisting of links $<ER_S,ER_H>$ where both ER_S and ER_H are N-labeled and ER_H was formed at the same time as$<ER_S,ER_H>$ – are therefore immune from link deficiency. Since it is causal chains of this kind that are the ultimate target of our reconstruction, link deficiency will be ignored in the remainder of the paper.

[27]For some discussion see (Kamp 2015).

[28]For the most part misinterpretations of names don't seem to do much damage. But to my knowledge the mechanisms that immunize name uses against occasional misunderstandings of this kind aren't very well understood and haven't yet been seriously explored.

By setting deficient links aside we finesse much of the structural complexity of ER networks. But there is another source of complexity to network structure that we won't be able to do without. Each link $<ER_S,ER_H>$ is created at some particular time, by the interpretation of some act of reference. These 'time stamps' of links impose a temporal order on them. Since this temporal order is important for the definition of causal chains, we make it part of the relational structure that will serve as the formal basis for our further considerations. That is, we define ER networks as sets of triples $<ER_S,ER_H,t>$, where t is a moment of time – I assume that time is like the real number structure $\mathcal{R}$, but this is of no great importance here – and ER_S and ER_H are linked by a vicarious anchor in the anchor set of ER_H that becomes part of this anchor set at time t. Let us assume moreover that we are looking at the uses of a name N for an entity $\mathbf{r}$ in a speech community SC over some period of time T (some finite interval of $\mathcal{R}$) and that T includes the baptism event in which N was introduced as name for $\mathbf{r}$. Note that for certain names, for instance the biblical names *Moses* and *Jonah*, but also probably even more ancient names like *Gilgamesh* or *Djoser*, T has to extend quite far into the past. (I am taking it for granted that these are all names of real people.)

The ER Networks defined in (24) are those whose ER links are the result of communication events in the course of T in which agents from SC add vicarious anchors to their N-labeled ERs for $\mathbf{r}$ as part of their interpretations of tokens of N. Among these events we also include those where a member of SC encounters a reference to $\mathbf{r}$ by a token of N in a text. For such cases I am making the somewhat simplifying assumption that the link triple $<ER_S,ER_H,t>$ has as first constituent the ER that enabled the author S of the text to use N to refer to $\mathbf{r}$; the second constituent is (obviously) the ER is ER_H of the reader H to which the new vicarious anchor is added; and t is the time at which H is reading this occurrence of N.[29]

Here is the definition of the notion of an N-labeled ER Network.

(24) (Definition of N-labeled ER Networks)

The *N-labeled ER Network for N as name for $\mathbf{r}$ within the community*

[29]The reason for wanting to include such links was already mentioned in footnote 21 of Section 3.4: in literate societies like ours encounters with names in texts are often the main source of name transfers, especially to members of the speech community who are born a long time after the name was introduced. Our reliance on texts to acquire command of old names becomes crucially important when their referents no longer exist. They are our remaining life lines to those referents. (Recall the remark about *Lavoisier* in footnote 21.)

SC *over the period of time* T, NW(SC,T,N,**r**), is the set of all triples $<ER_S,ER_H,t>$, where t in T, S and H belong to SC, ER_S and ER_H are both N-labeled ERs representing **r** and ER_H contains a vicarious anchor that is introduced into it at t as a record for the given use by S of N to refer to **r**.

When the members of SC behave throughout T in accordance with the rules of the language, including those for the processing of referential uses of proper names, then every member S of SC with an N-labeled ER ER_S that represents **r** will be connected by a 'causal' chain of elements of this Network to some member of SC that was present at the baptismal event which conferred N upon **r**. Formally:

(25) (Definition of causal chains)

A *causal chain of* the Network NW(SC,T,N,**r**) *that accounts for the command of* S *at* t *of* N *as name for* **r** is a finite sequence $<cl_1$, ..., cl_{n-1}, $cl_n>$ of some 'chain length' n, where (i) for each i = 1,..,n, cl_i is a link $<ER_{S_{i-1}},ER_{S_i},t_i>$ from NW(SC,T,N,**r**), (ii) S_n = S, (iii) for i = 1,..,n-1, $t_i \prec t_{i+1}$, (iv) $t_n \prec t$ and (v) S_0 was participant in the baptism in which **r** received the name N.

Definition (25) is the formal reconstruction I offer for the chains of *Naming and Necessity*'s causal chain account of names. Put in less technical terms:

1. A *causal chain accounting for the command at* time t within T *by* a speaker S who belongs to the speech community SC *of* the name N *as name for* **r** is a linear substructure of NW(SC,T,N,**r**) that is a causal chain according to Definition (25) above.

2. In a community SC in which everyone abides by the production and interpretation rules governing the referential use of names it will be the case for any member S of SC who at time t is in command of N as name for **r** that there exists a causal chain which accounts for S's command at t of N as name of **r**.[30]

[30] In (Almog, Nichols & Pepp 2015) the authors introduce the notion of a *perceptual chain*; they illustrate the notion with an example from (Kripke 1979*b*), to which they refer as the 'Smith-Jones example': S and H are out on a walk when S notes someone in the distance who is raking the leaves and whom she takes to be Jones. This prompts her to say to H: "Jones is raking the leaves." However, the person she is looking at isn't Jones, but someone else who goes by the name of Smith. The authors describe what is going on in this case as the 'intersection' of two perceptual chains that S is involved in, a short chain that leads from Smith to S via S's visual perception of him and a 'historical chain'

Definition (25) raises some further questions. For one thing, there will typically be many causal chains in a Network that connect a speaker S with someone present at the baptism. When all that we are concerned with is to account for the command of names in terms of the existence of a causal chain, then Definition (25) gives us all we want. But in general the network will contain many different causal chains accounting for S's command of N for $\mathbf{r}$ at t and the question could be raised whether some of those chains should be considered better accounts of S's command than others. Let CH(S,N,$\mathbf{r}$) be the set of all chains in the Network NW(SC,T,N,$\mathbf{r}$) that connect S's N-labeled ER for $\mathbf{r}$ with someone present at the baptism.[31]

---

that connects S with with the baptism of Jones and that, I take it, is to account for her command of the name *Jones* as name for Jones. If I understand the proposal correctly, then chain intersection would also have occurred if S had recognized the man she is looking at correctly as Smith, but in this case this would have involved the intersection of the perceptual chain that connects S with the man she is seeing and a different historical chain, the one that connects her with the baptism that conferred the name *Smith* on Smith. The analysis of perceptual recognition in terms of chain intersection says a little more about what is involved in such recognitions than can be found in the present paper (which has nothing to say about this at all). But as far as I can see, it doesn't tell us much about the historical chains that account for how speakers become competent users of names.

The authors of (Almog et al. 2015) seem to think that historical chains too are perceptual chains. I am not sure to what extent that bears on the question whether their version of the chain theory and the one attempted in the present paper are compatible. The answer, I presume, will depend on what we want to say about the interpretation events that are the origins of vicarious anchors. Are those events to be classified as perceptions? Perhaps. After all they are events in which the interpreter hears or reads a token of a proper name, so has an auditory or visual experience of the name. But even if utterance interpretations qualify as perceptions, the crucial difference drawn in the present paper – that between perceptions of entities and perceptions of their names, and derivatively that between vicarious anchors and the perceptual anchors of Section 3.3 – remains. It is the vicarious anchors that are responsible for the causal chains of Definition (25). Entity perceptions may play a part in network and chain formation too, for instance when a speaker makes an *introductory* use of a name in the sense of (Kamp 2015). But as far as I can tell, introductory uses of names play no part in *Naming and Necessity*, so in the present paper I have set those aside. Since I am unsure whether the distinction between vicarious and non-vicarious information is made in (Almog et al. 2015), it is not clear to me to what extent the historical chains of that paper and the causal chains of Def. (25) are alike.

[31] I am speaking here of 'S's N-labeled ER for $\mathbf{r}$', suggesting that there is just one such. But is that justified? Not when we allow for cases like Peter of Kripe's 'Puzzle about Belief', who has two *Paderewski*-labeled ERs for the scion of early 20-th century Polish destiny. Should the *Paderewski*-labeled ERs of someone like Peter be excluded from links in causal chains that account for his command of *Paderewski* (as name of this Paderewski)? I have no definite view of the matter and leave it as food for the *Paderewski* experts.

Which elements of CH(S,N,**r**) should be considered good candidates for the explanation of S's command of N as name for **r**? Here are some some candidates. One is the chain that optimizes for first exposures: Consider all links in the network of the form $<ER_{S'},ER_S,t>$, where ER_S is S's N-labeled ER for **r** and $ER_{S'}$ is an N-labeled ER for **r** of some other member S' of SC. Among these triples there is one for which its time t is earlier than the times of all the others.[32] If S' was a baptism participant, then we are done. (This is a very short chain!) If not, then consider all links $<\mathrm{ER}_{S''},\mathrm{ER}_{S'},t'>$ with $t' \prec t$ – again this set cannot be empty, for otherwise S' could not have the N-labeled Entity Representation $\mathrm{ER}_{S'}$ – and take the one with the earliest time from this set; and so on till a baptism participant has been reached.

But there are also other ways to single out plausible candidates. For instance, there are the shortest chain or chains in CH(S,N,**r**).[33] I cannot think of compelling reasons why either of these two types of chain – or any other type for that matter – should deserve preferential treatment in explanations of the command that speakers have of names. And in fact, can we think of reasons for wanting to disregard any chains from CH(S,N,**r**) in an account of S's command of N? Hardly. For one thing, full command of a name often requires, we noted, more than a single exposure to it. From that point of view the more chains, the better. And a furthr consideration in this connection: even after S has achieved what might be considered full command of N, additional exposures to uses of N by others, and arguably also S's own uses of N, may have the effect of reinforcing her command of N and strengthening her connection to **r**. I want to suggest therefore that we see CH(S,N,**r**) as the most natural way of accounting for S's command of N: It gets under way the moment the set CH(S,N,**r**) becomes non-empty. And as time goes on and CH(S,N,**r**) grows bigger, it too will grow, through the new chains that connect S with **r** that result from her further exposures to uses of N, but also, and perhaps even more so, through the new information about

[32] I am making the assumption here that no two vicarious anchors can be formed and added to their ERs by the same agent at exactly the same time.

[33] Note that such chains need not be the same as the 'earliest exposures' chain described in the last paragraph. To see this suppose for instance that S acquires trhe name *Wolfgang Amadeus Mozart* through countless mentions of it by others – people she talks to, radio announcers and so on – who also got the name from such mentions by yet others, and then lays her hands on Von Nissen's Mozart biography, written by a man who didn't personally know Mozart himself, but who intimately knew Mozart's wife Constanze. (Von Nissen was her second husband.) Since Constanze knew Mozart's father, undoubtedly a participant in Mozart's baptism, this gives a chain of length 3. It is a reasonable assumption that S's earlier exposures to Mozart's name all connect her to Mozart's baptism by chains that are much longer.

r that she acquires from those exposures.[34]

The chain sets CH(S,N,**r**) for various combinations of agents S, names N and referents **r**, may differ considerably in their internal structure. What structural properties are of interest to you will depend on your dispositions, scientific, philosophical or otherwise. Here is one such property, which might have some interest from a historiographic perspective. Some of the chain sets CH(S,N,**r**) that connect speakers S today with persons and other sorts of entities from antiquity have a structure that is reminiscent of the shape of an hour glass. The narrow hole in the middle of the glass is a single text in which N occurs. This text was written at a time when the referent **r** of N no longer existed. But let's assume that the author, who could not have known **r** directly, was nevertheless multiply connected with **r** via one or more people who participated in the baptism of **r** with N. The text then went dormant for many years and **r** completely disappeared from collective memory. When eventually the text was rediscovered, at which time it was the only extant source of N, the name N, and **r** with it, returned to public awareness, with growing numbers of people using N as name for **r**. From that point on speakers with a command of N will be connected with **r** by many chains, but all those chains will pass through the one text that has been rediscovered. Because of that any such speaker will feel her connection with **r** a precarious one, no matter how many chains there are in her chain set CH(S,N,**r**). Finding another text with occurrences of N that can be confidently identified as coreferential with the ocurrences of N in the rediscovered text would make a big difference if the texts could be shown to be independent of each other in the sense that neither author has been relying on the other.[35]

So much for the structure of chain sets that connect speakers with the origins of names for entities from a distant past. There are other kinds of questions about ER Networks that can be asked as well. We can, instead of looking at such Networks from the perspective of some particular speaker S, also look at them from that of a given name N, either as name of some particular referent **r**, or as a name that is used in the speech community as the name of multiple referents. And lastly, we can look at Networks from the perspective of an entity **r** that is represented by ERs occurring in Networks for different

[34]Compare the discussion of multiple exposures to name uses in (Evans & Altham 1973).

[35]To be precise: suppose that N occurs both in text T_1 from author A_1 and in text T_2 from author A_2. We say that A_1 *relies in her use of* N *in* T_1 *on* author A_2 iff every chain in CH(A_1,N,**r**) contains a link of the form $<ER_{A_2},ER_B,t>$ (for some B or other). T_1 and T_2 were independent with regard to their occurrences of N iff neither A_1 relied in her use of N in T_1 on A_2 nor A_2 on A_1.

names (all of which have been used, by some people, at some times within T, to refer to **r**). Let me end this section with a couple of questions that could be asked from these perspectives.

The first has to do with the ambiguity of names. The Network we need to look at in this case is the union of the Networks NW(SC,T,N,**r**) for a fixed name N, but different entities **r** for which N has been used as a name within SC in the course of T. Suppose we subdivide T into k equal portions T_1,.., T_k. For each period T_i (i = 1,..,k) there is a certain set Ref(N,T_i) of entities for which N has been used as a name within SC during T_i. These sets give us a window on the 'popularity' of N during T_i; by juxtaposing these sets for the different T_i we get something like a popularity profile for the name N over the entire period T.

When we look at ER Networks from the perspective of different names for the same entity (as opposed to different entities that have been referred to by the same name), other questions come to mind: Focus on some entity **r** and consider the union of the Networks NW(SC,T,N,**r**) for the different names N that are or have been in use for **r**.) Assume that T starts when **r** comes into being (e.g. when **r** is born, in case **r** is a person) and assume again that T is subdivided into k equal portions T_1,..,T_k. For each period T_i we can now ask who in SC knew **r** under some name or other during T_i. In this way we get some kind of 'popularity profile' for **r**. What might we be able to learn from such profiles, or from the structure of the Networks in terms of which they are derived?

There are many more such questions that can be asked about networks and chains. But the three I have mentioned should have given enough of the general flavor and I won't embroider any further. But there is a further question I want to raise at this point. It is a kind of second order question: Some questions about network or chain structure might be of some interest from some perspective. But is there any *philosophical substance* there might be to them? I suspect that the prospects are not good, and if I have included a couple of them here, it has been with the awareness that they will seem oddly placed in an essay that is intended as a tribute to Saul Kripke. If there is any justification at all for mentioning such questions here, it is that I see it as a good way of bringing out to what extent and in what way the Causal Theory of Names is a social construct. That causal chains are social constructs may be obvious enough, irrespective of any formal reconstruction of them. But the reconstruction I have proposed shows how inseparable this social dimension is from an issue that is usually treated as belonging somewhere

quite else: How is linguistic meaning determined by linguistic form? In the next and final section we will have a quick look at the implications of this inseparability for the methodology of natural language semantics.

5 Names and the Social Dimension of Meaning

Central to our reconstruction of the Causal Theory of Names is the way coreference links are established between Entity Representations of different agents. According to the story told in Section 3.4 these links arise as a kind of secondary effect in the course of utterance interpretation – utterance recipients need them for the semantic representations they construct. But the very connections that interpreters must assume to get their interpretations right create the linguistic cohesion within the speech communities to which they belong. Intersubjective linking is an integral part of utterance interpretation.

If this is right – and I hope that the arguments of this paper have made the case that it is – then the role of the interpreter, who can acquire the use of names by the very same procedures that enable him to understand the messages conveyed by sentences containing names, must be at the center of an account of the truth-conditional content of such sentences. And that poses a problem for a conception of natural language semantics that has been dominant since the beginnings of formal semantics more than fifty years ago: The principal task for natural language semantics is to formulate accounts of the truth conditions of sentences (and perhaps also larger units, like dialogues or texts) and these truth conditions should be treated as properties that natural languages have qua autonomous systems, which are what they are independently of how and by whom they are used. In semantic theories that conform to this methodology, truth conditions must be described without any reference to their users.[36]

This methodological principle has been remarkably successful; most of the progress in formal semantics of natural languages over the past half century has been made in work that has been guided by it. There is nothing surprising

[36]This doesn't mean that truth conditions must be treated as independent of the context of use. For instance, as early as in the work of Montague, the starting point and paradigm for the conception of formal semantics as a science of natural languages treated as user-independent systems, there is a clear awareness of the context dependence of indexical expressions like *I*, *you*, *now*, *here*, *tomorrow* (Montague 1970*b*), (Kaplan 1989). What the conception excludes are references to the mental states of interpreter and speaker.

about this. Human languages show a remarkable degree of community-wide uniformity and stability, also and in particular in how the forms of their expressions determine their meanings. If it wasn't for such uniformity and stability, our languages couldn't be the effective toolboxes they are for expressing and communicating the often complex information that we want to make explicit and get across to others. Semantic theories that describe the form-meaning relation of human languages without explicit reference to the mental states of users work as well as they do because the user can normally rely on this uniformity and stability.

But where do the uniformity and stability of the form-meaning relations of human languages come from? What brings them about and maintains them? That is a question a general theory of linguistic meaning should be allowed to ask too. And here too we can find the beginning of an answer in *Naming and Necessity*. The reconstruction of the causal chain account I have outlined in this paper helps, I hope, to make the nature of this answer a little clearer: Interpreters of name tokens follow a strategy that aligns their command of those names with that of the producers of those tokens. It is a tempting thought that this strategy, which seems to cater so well to the needs of a community that must be able to rely on the uniformity and stability of its language, is operative not only in connection with proper names but more generally. But is that so? That it is – to some extent at least – is something that we can also learn from *Naming and Necessity*. Natural kind terms – words like *tiger*, *lemon*, *gold*, *water*, *carbon dioxyde* – refer to kinds in much the same way that proper names like *Julia* or *Paris* refer to their bearers. There are important differences between the two cases. Kind nouns differ from proper names in that they are more commonly used to make statements about members of the kinds they refer to than about those kinds qua kinds.[37] But the similarities – how proper names and kind terms

[37]Another important source of the view of kind terms is (Putnam 1975). Putnam discusses at length how the meaning/use of kind terms is explained to those who do not yet know them. One way to do this is by making an 'ostensive' use of the term. For example, to explain the meaning of *lemon* to someone one can say 'This (here) is a lemon' while pointing at some lemon, as an instance of the kind that *lemon* refers to.
Note well, however, that the ostensive use of kind terms are not the only use that enables novices to learn their meanings. Natural kind terms can also be inferred from their 'standard' uses. An example of the standard use of a kind term is that of *lemon* in the statement 'This lemon is spoiled', whose purpose is to inform the addressee about some particular property of some particular lemon. Ostensive uses of natural kind terms are like introductory uses of proper names, as when I say to you: "This is Julia", gesturing towards my girlfriend, who is standing next to me. Such uses of proper names are clearly different from the standard uses on which I have concentrated in the present paper. Os-

are acquired and how their reference is stabilized by their use – are striking nonetheless.[38]

Does the mechanism that is operative in our use of proper names and natural kind terms cover yet other domains as well? As things stand, I do not dare to say. But I think there can be no doubt that different mechanisms are at work elsewhere which are like this one in that they too create uniformity and stability and enable language users to 'learn while doing'. Here is one example: relative gradable adjectives, such as *tall* or *expensive*. In order that the positive form of such an adjective A can be truly predicated of an entity $\mathbf{r}$ that is an instance of a comparison class defined by a noun N – that is: in order for the statement "$\mathbf{r}$ is an A N" to be true, given that $\mathbf{r}$ is an N – the degree to which $\mathbf{r}$ satisfies A must exceed a certain 'threshold degree' for A and this comparison class (Kennedy 2007). This entails that to understand what a speaker who is making a statement of this form is saying, the interpreter must know what this threshold is. For instance, suppose you say (26) to me. If I am to understand what you are telling me about John's height I must know what threshold you are assuming for tallness among basketball players.

(26) John is a tall basketball player

But what determines such thresholds? It has long been assumed that thresholds for gradable adjectives are at best loosely determined. In fact, the positive forms of gradable adjectives have been treated as paradigms of vagueness, and for an adjective like *tall* its vagueness *is* threshold vagueness, vagueness about where the thresholds for different comparison classes are situated along the adjective's scale.

Over the past decade attention has turned increasingly to the use-based mechanisms that calibrate the likely positions of these thresholds within speech communities, making use of the fact that interpreters of statements like (26) sometimes have independent information about how tall the subject is. (If I know that John's height is six feet seven inches, then I can infer from (26) that you assume a standard for tallness among basketball players that is less than six feet seven inches.) Currently there are several models on the

tensive uses of kind terms differ from their standard uses analogously. For discussion of standard and introductory uses of proper names see (Kamp 2015).

[38]So much so, in fact, that it is easy to extend MSDRT with *Kind term-labeled Kind Representations*. To my knowledge this extension hasn't been carried out in print, but it is fairly clear how it should go.

table for how inferences of this kind can calibrate the thresholds that are assumed within a speech community (Potts n.d.), (Lassiter & Goodman 2013), (Qing & Franke 2014). These models all treat semantic information as probability distributions, and assume a radically different framework for doing natural language semantics than the one I have tacitly presupposed in this paper. That in itself is a *prima facie* reason to expect that the process we have focused on in this paper and the processes of adjective use described by those Bayesean models have little in common apart from the fact that they both make coordination and calibration inevitable concomitants of interpretation. When we look at these Bayesean models more closely, we find that expectation confirmed.

The second example has to do with the emergence of 'non-literal' uses of predicate words (nouns, verbs, adjectives, prepositions). Well-worn examples are *shark* in "My lawyer is a shark", *bark* for what officers do when they shout instructions at their men, or idioms like *bite the bullet*. It can be assumed that the first instances of such uses were like new live metaphors, which contributed to the sentences containing them not only the criteria for predicate satisfaction that remain after they have become conventionalized items of the lexicon, but also the surplus of a surprising analogy with the pre-existing 'literal' use of the word or phrase. But when conventionalization sets in, the surplus loses its surprise value and gradually fades; and often it fades away completely and the connection with the literal use is lost altogether. (Who today is still aware that our current use of *bite the bullet* as a way to refer to decisions to go ahead with something that one knows is going to be painful or unpleasant originated as a simile with a person who has to undergo surgery without anesthetics and who is given a lead bullet to bite on it to cope with the pain as well as he can?)

Three observations about the emergence and spread of new word meanings:

1. Processes that start when a word is first used with a new meaning, which can be understood because of a salient relationship with its established ('literal') meaning, but where the new meaning then gradually becomes self-supporting and the new meaning becomes a separate item in the lexicon, are exceedingly common and probably the most important force in the extensions and refinements of the vocabularies of human languages in the course of their histories.

2. At least the early stages of these processes look like instances of the general pattern of 'learning while doing' and community-wide coordination:

Interpreters can catch on to the new meaning because of its ties to the old meaning they are familiar with, helped in this by the cultural context they share with the speaker or author, and what the speaker or author is likely to want to say with her utterance. Thus the new meaning gets transferred to him, and with the usual calibrating effects. (The later development stages of non-literal meanings seem different. The new meaning has now become emancipated from the literal meaning and the word as denotation of its non-literal meaning is now one of the many lexical predicates of the language, with its own conventionally fixed application criteria. The general question how predicate words can be learned and their application criteria kept stable is one of the central aims of the study of linguistic meaning from the perspective we are considering. A detailed account of what happens in the early stages of non-literal meaning development can, I believe, be an important step towards a better understanding of this more general problem.)

3. In the psychological and psycholinguistic literature detailed descriptions can be found of the mechanisms that lead from imaginative uses that can be made of words that endow them with new meanings to the emancipation of those meanings as independent predicate meanings in their own right. (Gentner & Grudin 1985). Yet, much of what is relevant to the study of these mechanisms from the perspective on language discussed in this section doesn't seem to be very well understood. However, in recent years more formal methods have been brought to bear on the study of these mechanisms, especially in computational linguistics. Here too new research paradigms have been getting under way in which there is a strong emphasis on the social dimension of meaning.

The two mechanisms I have touched upon in this section – threshold determination for relative adjectives and the birth and growth to adulthood of non-literal meanings – are, I repeat, like the mechanism that governs the use, acquisition and community-wide stability of proper names, which has been the central focus of this paper. All three work in the way that should be expected for mechanisms that are part for a highly complex and conventionalized social practice such as the human use of language. The recent explorations of these mechanisms that I mentioned in passing in this section are in the spirit of this perspective and I am confident that the trend they have been setting will steadily gain in scope and importance. But if detailed work of this kind is a recent phenomenon, the motivation for it has long been there for all of us to see. And perhaps the most prominent place where it has been clearly visible for all this time is *Naming and Necessity*. What *Naming and Necessity* taught us about the semantics and pragmatics of proper names

was a radical innovation at the time, and that was understood pretty well from the start. What could not be seen then, but what we can see today, is that it was also, in its social implications, prophetic.

References

Almog, J., Nichols, P. & Pepp, J. (2015), A unified treatment of (pro-)nominals in ordinary english, *in* A. Bianchi, ed., 'On Reference', Oxford University Press.

Carnap, R. (1947), *Meaning and Necessity*, University of Chicago Press, Chicago.

Cumming, S. (2019), 'Names', *Stanford Encyclopedia of Philosophy* .

Davidson, D. (1963), 'Actions, reasons and causes', *The journal of Philosophy* **60 (23)**, 685–700.

Evans, J. & Altham, J. (1973), 'The causal theory of names', *Proceedings of the Aristotelian Society, Supplementary Volumes* **47**, 187–225.

Farkas, D. (2002), 'Specificity distinctions', *Journal of Semantics* **19**, 1–31.

Geach, P. (1962), *Reference and Generality : An Examination of Some Medieval and Modern Theories*, Cornell University Press, Ithaca.

Gentner, D. & Grudin, J. (1985), 'The evolution of mental metaphors in psychology', *American Psychologist* **40**, 181–192.

Geurts, B. (1999), *Presuppositions and Pronouns*, Elsevier Science.

Heim, I. (1982,1988), *The Semantics of Definite and Indefinite Noun Phrases*, Garland Press.

Heim, I. (1992), 'Presupposition projection and the semantics of the attitude verbs', *Journal of Semantics* **9**(3), 183–221. Special Issue: Presupposition, Part 1.

Hintikka, J. (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca N.Y.

Kamp, H. (1981), '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. (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. (2021), Sharing real and fictional reference, *in* E. Maier & A. Stokke, eds, 'The Language of Fiction', Oxford University Press.

Kamp, H. (n.d.*a*), An informal introduction the conceptual foundations, architecture and applications of mental state discourse representation theory.

Kamp, H. (n.d.*b*), Msdrt: motivations. formal definitions and an application.

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., 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.

Kaplan, D. (1989), Demonstratives, *in* H. W. J. Almog, J Perry, ed., 'Themes from Kaplan', Oxford University Press.

Kennedy, C. (2007), 'Vagueness and grammar: The semantics of relative and absolute gradable adjectives', *Linguistics and Philosophy* **30**, 1–45.

Kripke, S. (1979*a*), A puzzle about belief, *in* A. Margalit, ed., 'Meaning and Use', Reidel, Dordrecht.

Kripke, S. (1979*b*), Speaker's reference and semantic reference, *in* P. French, T. Uehling & H. Wettstein, eds, 'Contemporary perspectives in the philosophy of language', The University of Minnesota Press.

Lassiter, D. & Goodman, N. (2013), Context, scale strucrture, and statistics in the interpretation of positive-form adjectives, *in* 'Proceedings of SALT 23', pp. 587–610.

Maier, E. (2016), 'Attitudes and mental files in discourse representation theory', *Review of Philosophy and Psychology* **7(2)**, 473–490.

Montague, R. (1970*a*), English as a formal language, *in* B. V. e.a., ed., 'Linguaggi nella Societa e nella Tecnica', Edizioni di Communita, pp. 189–224.

Montague, R. (1970*b*), 'Universal grammar', *Theoria* **36**, 373–398. Reprinted in: R. Thomason (ed), Formal philosophy; selected papers of Richard Montague, Yale University Press, 1974.

Morris, C. W. (1946), *Signs, Language and Behavior*, Oxford University Press.

Partee, B. (1984), 'Nominal and temporal anaphora', *Linguistics and Philosophy* **7**.

Perry, J. (1980), 'A problem about continued belief', *Pacific Philosophical Quarterly* **61**, 317–332.

Potts, C. (n.d.), Interpretive economy, schelling points, and evolutionary stability.

Putnam, H. (1975), The meaning of "meaning", *in* 'Language, mind, and knowledge', Minneapolis: University of Minnesota Press.

Qing, C. & Franke, M. (2014), Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model, *in* 'Proceedings of SALT 24', pp. 23–41.

Recanati, F. (2012), *Mental Files*, Oxford.

Van Der Sandt, R. (1992), 'Presupposition projection as anaphora resolution', *Journal of Semantics* **9**, 333–377. Special Issue: Presupposition, Part 2.