Why my *I* is your *you*: On the communication of *de se* attitudes*

Emar Maier
University of Groningen

Abstract The communication of *de se* attitudes poses a problem for "participant-neutral" analyses of communication in terms of propositions expressed or proposed updates to the common ground: when you tell me "I am an idiot", you express a first person *de se* attitude, but as a result I form a different, second person attitude, viz. that you are an idiot. I argue that when we take seriously the asymmetry between speaker and hearer in semantics this problem disappears. To prove this I propose a concrete model of communication as the transmission of information from the speaker's mental state to the hearer's. My analysis is couched in Discourse Representation Theory, a formal semantic framework that linguists use for modeling conversational common ground updates, but that can also be applied to describe the individual speech participants' dynamically changing mental states.

Keywords: Attitudes *De Se*, Communication, Mental Files, Discourse Representation Theory, Dynamic Semantics, Presupposition, Indexicals

Contents

* This is the second version of this manuscript, uploaded September 12, 2014. I thank an anonymous reviewer and editors Stephan Torre and Manuel García-Carpintero for constructive feedback. Further thanks to Hans Kamp, Julie Hunter, and François Recanati for fruitful discussions abour these topics. I also thank the organizers and audience of the workshop *Centered Content and Communication* in Barcelona, May 2012. This research is supported by the EU under FP7, ERC Starting Grant 263890-BLENDS.

1 Introduction

The traditional account of communication in terms of propositions runs as follows. Mary believes that linguistics is hard, i.e. she stands in the belief relation to the proposition that linguistics is hard. She wants to communicate this belief to John. To do so, she produces an English sentence that expresses the believed proposition, e.g. "Linguistics is hard". John hears the sentence and interprets it as expressing the proposition that linguistics is hard. If John has no reason to distrust Mary in this matter he will then add this proposition to his set of believed propositions. In other words, a proposition has been transmitted from Mary's beliefs to John's via a linguistic expression that encodes it.

This general view of communication is compatible with different notions of proposition and belief. We may think of propositions here as Lockean Ideas, Fregean Thoughts, or sets of possible worlds. In this paper I take the latter conception as my point of departure.

However, as Lewis (1979) has shown, not all beliefs correspond to possible worlds propositions. Some thoughts require a more fine-grained notion of content, such as self-ascribed properties or, equivalently, centered propositions. Such beliefs are known as *de se* beliefs. Unfortunately, as Stalnaker (1981) observes, the simple picture of communication sketched above does not extend from propositional to *de se* belief (cf. Ninan 2010).[1] Consider

1 The problem goes back to Frege (1918), who presents a slightly different diagnosis: there is a special, first person sense of *I am wounded* that cannot be communicated at all, cf. (Recanati 2012: VIII). Following Stalnaker, I take it as a given that first person *de se* thoughts *can* be

the case of Lingens, an amnesiac lost in the Stanford Library, who wants to communicate his belief that he himself is lost, i.e., in Lewisian terminology, his self-ascription of the property of being lost. The obvious way for Lingens to express this belief to the librarian would be to use an indexical and say "I am lost". And indeed, if the librarian hears Lingens utter that sentence, he can interpret it as meaning that he, the person addressing him, is lost and consequently help him out. Now note that what the librarian comes to believe in this way is not the same self-ascribed property that Lingens set out to express. Lingens self-ascribes the property of being lost, expresses that by saying "I am lost", and as a result the librarian self-ascribes the property of being addressed by someone who is lost. The question now is, how exactly did Lingens's first person belief, expressed with a first person pronoun, turn into a non-first person belief when it reached the librarian?

To answer this question I propose a formal model of communication that clearly distinguishes the speaker's production perspective from the hearer's interpretation perspective. This requires first of all an explicit model of the speech participants' mental states, paying particular attention to *de se* beliefs (section 2). The second ingredient is a "participant-neutral" theory of linguistic communication as dynamic common ground updates (section 3). In section 4 I combine these two independently motivated theories into an asymmetric model of communication, clearly distinguishing the speaker's production perspective from the hearer's interpretation perspective. Focusing on the first and second person I demonstrate in section 5 how the model deals with the transmission of *de se* beliefs via indexicals.

2 Representing mental states

2.1 Discourse Representation Theory

Today, Discourse Representation Theory (DRT, Kamp & Reyle 1993) is typically presented as a specific type of formal semantics, well suited for dealing with semantics/pragmatics interface phenomena like presupposition and anaphora resolution. DRT's formal language of Discourse Representation Structures (DRS) is used to represent the common ground between speaker and hearer. The explanatory power of the framework lies in the algorithms for updating these common ground representations in response to linguistic utterances. In section 3 I present this DRT model of common ground updates in some detail.

communicated – the only question is how.

What is often glossed over is Kamp's (1981) original motivation of reconciling Fregean formal semantics (as championed at the time by Montague (1973)) with a traditional, Lockean cognitive theory of communication in terms of speakers' and hearers' mental states. To this end, Kamp in his original presentations describes DRSs as representations of the mental state of the hearer, rather than of the more abstract notion of a common ground. What sets this cognitive conception of DRT apart from purely cognitive theories like Fauconnier's (1994), is that the DRS language has a model-theoretic interpretation, much like that of (intensional) first-order logic. Hence, in addition to its cognitive interpretation, a DRS also represents the actual truth conditions of a sentence or discourse.

Linguists have since stripped DRT of its cognitive interpretation. But Kamp and a few others have kept it alive, even extending DRT to a representational theory of attitudes (Kamp 1990, Asher 1986). In the remainder of this section I present a novel version of such a DRT-based theory of mental states.

2.2 Mental files, anchors, and attitudes

The starting point of Kamp's general framework for describing mental states (Kamp 1990, Kamp et al. 2003, Kamp 2011) is that mental states are (i) compartmentalized into beliefs, desires, fears, intentions, etc., and (ii) these compartments are highly interconnected. For instance, my mental state could contain the belief that there's a monster under my bed and, dependent on that belief, the hope that *it* won't wake up. This dependence is cashed out in the same way as anaphoric dependencies in discourse are modeled in standard DRT, viz. by sharing accessible discourse referents.

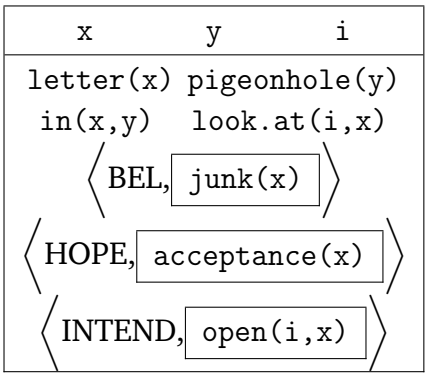
To model singular attitudes, Kamp further introduces the notion of "entity representations" or "internal anchors". These correspond rather closely to what philosophers have called dossiers (Grice 1969), or mental files (e.g. Perry 1980, Recanati 2012). I will use the latter term. Mental files contain the descriptive content we've obtained about the actual world through acquaintance with particular objects. A mental file thus serves as a cognitive mode of presentation of the object that is the causal source of the information stored in it.[2]

For concreteness, the box below represents the mental state I'm in when I

2 As Ninan (2014) observes, mental files are sometimes described as syntactic objects, containing predicates, and sometimes as semantic objects, containing information. In the current proposal this mixing of metaphors is indeed harmless because in section 2.4 I explicitly provide a model-theoretic interpretation, mapping the syntactic files to the corresponding semantic objects.

discover a letter in my pigeon hole, believing it's publisher's junk mail, but hoping it's an acceptance letter for a recent grant application, and intending to open it right away.

(1)



This mental state description contains at the global level three mental files (representations of the letter, the pigeon hole, and myself) and on top of that three attitude descriptions (a belief, a hope, and an intention). The attitudes are represented as DRS boxes labeled with a mode indicator (BEL, HOPE, INTEND). The underlying mental files are represented jointly by the global discourse referents (x, y, i) and the global conditions (letter(x), look.at(i,x), ...). Global discourse referents are accessible to the attitude descriptions, allowing us to represent different attitudes "about the same thing". In this case we see a belief, a hope and an intention that are all about the same letter, by virtue of sharing the discourse referent x. The global conditions specify my descriptive, cognitive modes of presentation of the letter, pigeon hole, and myself. This descriptive content derives from the ways I am acquainted with the actual causal sources of these files.

The current presentation departs from Kamp's and from the traditional manila file folder metaphor in taking seriously the idea that the contents of different mental files are often intertwined and cannot be neatly separated:[3] if I see Sue and Mary shaking hands, I could enter *shakes hands with Sue* into my Mary-file and *shakes hands with Mary* into my Sue-file, but since I really perceive just one hand-shaking event, it seems more natural to give up the boundaries between different file contents and just represent the "mental file

3 Cf. (Perry 2003: 53): "When it comes to forming a picture or battery of metaphors for how our minds handle relations, the file folder analogy begins to limp badly, and something along the lines of relational database theory would work better." Pryor (2013) works out an interesting alternative implementation of this idea of an interconnected web of mental files in terms of graphs, with nodes representing the files themselves and labeled edges representing the relations between them.

cabinet" as a whole.

Note that we can still distinguish and count individual mental files, simply by looking at the discourse referents. By way of illustration, a mental state description with two global discourse referents, x and y, and two global conditions x *is called London* and y *is called Londres*, represents a different mental state than one with only a single discourse referent x and the descriptions x *is called London* and *lx is called Londres*. In the former case, the subject has two files, representing the fact that she believes to be acquainted with two distinct cities; in the latter she believes to be acquainted with a single city that has two names. Accordingly, I will sometimes conveniently refer to files via their discourse referents, e.g. the attitude description in (1) contains the files x, y and i.

The actual causal sources of mental files are represented outside of the mental state description proper. Kamp formalizes the causal links between a mental state and its surroundings as a mapping from discourse referents to entities. In our example, this so-called "external anchor" maps the file x to the actual letter, y to the pigeon hole, and i to me:

(2)
$$\begin{bmatrix} \mathtt{x} \mapsto \text{letter} \\ \mathtt{y} \mapsto \text{pigeon-hole} \\ \mathtt{i} \mapsto \text{Emar} \end{bmatrix}$$

The external anchor allows us to capture singular attitudes: any attitude compartment that depends on a discourse referent introduced by an externally anchored file is a singular attitude, about the causal source of the file. In our example all three attitudes, the belief, the hope, and the intention, depend on the mental file x, representing the letter as something I am seeing in my pigeon hole. Consequently, they are *about* the causal source of that file, the actual letter.

Before I go into the formal semantics in section 2.4, let me further illustrate the representational framework by applying it to some slightly more interesting scenarios.

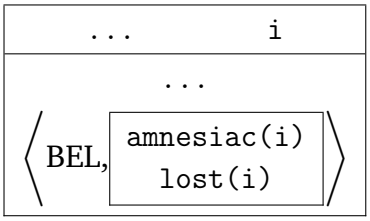
2.3 Three case studies

2.3.1 Attitudes *De Se*

This paper is about the transmission of *de se* beliefs, so we need a way to represent those. The idea is to use two dedicated indexical discourse referents i and n to represent the subject's first person *de se* center and subjective now, respectively. In (3) we see a fragment of the mental state of Lingens

who self-ascribes the property of being a lost amnesiac. Here and in the following I will focus exclusively on the person domain, ignoring tense and `n` for simplicity.

(3)



The self-file `i` is like other mental files in that it's accessible to all attitudes and other files. However, following Evans (1982) and others, I assume that subjects have a privileged, direct access to themselves, not mediated by descriptive modes of presentation. This means that the unlike regular object files, the self file need not contain any descriptive content representing relations of acquaintance.[4]

2.3.2 Double vision

The philosophical literature on attitudes and ascriptions is replete with puzzles where an agent forms two distinct representations of something, failing to realize that it is actually the same thing:[5] Frege's Babylonians see a distinct morning and evening star, Quine's Ralph sees mayor Ortcutt as distinct from the suspicious figure in the alley, Kripke's Pierre believes London is terrible,

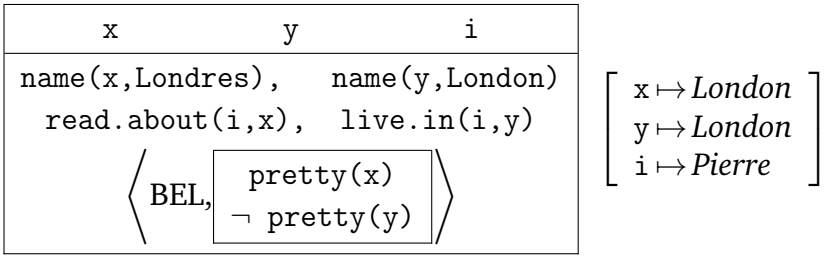
4 In other words, I'm following Lewis: we are acquainted with ourselves through the acquaintance relation of identity. In line with this admittedly controversial conception of the self, I will cash out the *de se* nature of attitudes involving `i` model-theoretically by stipulating that `i` evaluated with respect to a centered doxastic alternative always picks out the center (cf. section 2.4). García-Carpintero (p.c.) has objected that an alternative view where our acquaintance with the self is mediated by "bodily features of experiences, including the 'center-of-perspective' feature of visual experience" would fit as well or better with the general framework advocated here, and would better allow us to account for Immunity to Error through Misidentification, among other things. I will leave this for future research.

5 The opposite situation – one mental representation; two distinct objects – is discussed less frequently. As a reviewer points out, such a case is arguably problematic for a mental files framework like the one I'm developing here. Say, Mary is acquainted with both John and his twin brother without realizing she is dealing with two different people. It seems that she would have a single mental file anchored to two distinct individuals. To accommodate this technically we could give up the requirement that the external anchor is a function, but what would this mean conceptually? Are Mary's thoughts about John singular thoughts about two people? The easiest way to avoid this conclusion would be to assimilate such cases to the faulty perception cases discussed below in 2.3.3 below, i.e., rather than having two anchors, Mary's John file is in fact unanchored.

but Londres is pretty, Perry's shopper believes himself to be distinct from the shopper with the torn sack, etc. In the current framework, each of these cases involves a mental state description with two distinct files, each associated with different descriptive contents based on the different ways of being acquainted with an entity, but the external anchor maps both files to the same real-world entity, the actual causal source.

Pierre's predicament, for instance, can be represented as follows (Kripke 1979):

(4)



In Pierre's mental representation of the world there are two distinct cities: one is the city he read about in his childhood in France, called *Londres*; the other is the city he lives in, called *London*. Based on these two epistemic links, he has formed two files, and through these he can have singular beliefs, hopes etc.. We, as outside observers, know that his beliefs are in fact inconsistent. As represented in (4) by the external anchor, both mental files derive from a single source, so both beliefs are in fact about the same city. However, intuitively, the fact that Pierre believes the one city to be pretty and the other not does not entail that Pierre himself is irrational in the sense that his internal mental state is logically inconsistent.

To reconcile this apparent contradiction I will provide in 2.4 a model-theoretic semantics for mental state descriptions that defines both a narrow and a wide content of attitudes. The narrow content of Pierre's beliefs as represented in (4) should be computed on the basis of the descriptive conditions in the belief box and the mental files on which those depend. More specifically, the narrow belief content expressed by (4) should be the (centered) proposition that the city the subject knows as Londres is pretty and the city he knows as London is not. This gives us the non-contradictory interpretation that captures what goes on in Pierre's head. By contrast, the wide content of Pierre's beliefs is computed by evaluating the same belief conditions, but relative to the external anchor, bypassing the descriptive content in the mental files. Computing the wide content of (4) will give us the singular proposition about London that it is both pretty and not pretty, a genuine contradiction.

2.3.3 Faulty perception

So far we've discussed mental files based on actual acquaintance relations. Formally, every file we saw was externally anchored. Given that mental files are supposed to represent the objects of our *de re* attitudes, based on our acquaintance with our surroundings, this is as it should be. Hence, Kamp's (2011) slogan: "no internal anchor without an external anchor". But what if I merely hallucinated the letter in my pigeon hole? My narrow mental state in such a scenario will be the same as before, i.e., I cannot distinguish between the two situations. But now there is no causal source, i.e., no external anchor for the letter file. Consequently, the narrow contents of my beliefs, desires, hopes, remain the same while no wide content is expressed by my attitudes (that I take to be) about the letter.

In order to accommodate such cases, I follow a suggestion from Recanati (2012) to the effect that the external anchoring of all mental files constitutes a normative requirement: mental files should be, and hence can be expected to be, externally anchored. The agent, in any case, presumes all her mental files to be externally anchored. Thus, the mental file for the letter in my pigeon hole plays the same role inside my mental life regardless of whether it's properly anchored or hallucinated.

Kamp goes even further: if a mental file "has no external anchor corresponding to the representation's internal anchor (i.e. there is no entity to which agent and representation are causally related *in the way the internal anchor describes*), then the internal anchor is 'ungrounded'". (Kamp 2011: p.5, emphasis added) In other words, not only does a mental file require an external anchor, the descriptive content of the file needs to mirror precisely the actual causal relation between agent and *res*. Again, this is best thought of as a normative ideal that, in reality, is not always achieved. Again, the agent herself assumes all her mental files to be anchored to individuals that actually exemplify the properties associated with them in her mental files.[6]

More should be said about unanchored files. For instance, García-Carpintero (2010) argues that we can have singular thoughts about fictional entities like

6 At this point Kamp and I part ways with Recanati, who does not require the content of the file to match the actual relation of acquaintance. Part of the reason for the disagreement might be that Recanati does not distinguish mental file content from beliefs (and other attitudes), so his files must allow all kinds of information that an agent associates with a *res*, not just the relational acquaintance information (e.g. $\texttt{see}(\texttt{i},\texttt{x})$). My tentative suggestion is that relational acquaintance information about objects goes in the files and any additional information we learn or infer about objects generally goes in the belief box. In section 4 below I demonstrate for instance how information gathered through linguistic communication ends up in the belief box rather than in a file.

Sherlock Holmes. If so, we should allow mental files which even the agent herself assumes not to exist. For this purpose, Recanati actually introduces a special kind of files, *indexed files*, but this is beyond the scope of the current paper.

2.4 A model-theoretic interpretation

The line drawings above may give a pretty picture of, say, Pierre's mental state, but what does it really mean to say that Pierre has a mental state as described in (4)? In what sense does Pierre's mind contain such a DRS-like object?

As announced in section 1 I take as my point of departure the familiar possible worlds conception of propositions and beliefs. A person's beliefs are described by a set of possible worlds, her doxastic alternatives (Hintikka 1969). We can explicate the notion of a doxastic alternative as follows: w' is a doxastic alternative of a in w (notation: $w' \in Dox(a,w)$) means that if you take a from world w, freeze her mental state, and place her in world w', she will not be able to tell the difference. We then say that this person believes the proposition that it is raining iff all her doxastic alternatives are worlds where it is raining: a believes proposition $p(\subseteq W)$ in w iff $Dox(a,w) \subseteq p$

The Lewisian shift from propositions to properties as the objects of belief can be formalized as a shift from worlds and propositions to centered worlds (formalized as world–individual pairs) and centered propositions (i.e., sets of centered worlds), respectively. We take $Dox(a,w)$ to denote a set of centered worlds, i.e. $\langle w',a'\rangle \in Dox(a,w)$ means that if you place a in w' and let her experience it from the perspective of a', she will be unable to distinguish it from w as experienced from her own perspective. We then formalize *de se* belief as follows: Lingens self-ascribes in w the property of being lost iff $Dox(Lingens,w) \subseteq \{\langle w',a'\rangle | a' \text{ is lost in } w'\}$.

The same story applies to other attitudes: to model desires we have $Bul(a,w)$ denoting the set of a's centered buletic alternatives in w, and for imagination we have a set of imagination alternatives. A person's full mental state can thus be characterized as a sequence of subsets of C, the set of centered worlds. Formally, these attitude characterizations of people across possible worlds are part of the model, i.e., a model $\mathfrak{M}$ is a tuple $\langle D,W,I,\langle Dox,Hope,Bul,\ldots\rangle\rangle$ with $C = D \times W$, and $Dox,Hope,Bul,\ldots : C \to \mathscr{P}(C)$.

So how do we relate this Lewis/Hintikka-style set-theoretic conception of an agent's various attitudes, to our syntactic, DRT-based mental state descriptions? As a first approximation, the central definition runs as follows:

M is a partial description of a's mental state in w iff the narrow contents of the belief, hope, desire, etc. components within M are compatible with the sets of doxastic, hope, buletic, etc. alternatives of a in w. Making this precise, taking into account also the mental files in M, requires first a model-theoretic interpretation of the various labeled parts of a mental state description in terms of centered worlds.

The reader who sees that this can be done, and is not interested in the formal details, may safely skip the remainder of this section.

We start from the intensional interpretation of standard DRT (see, e.g. Geurts (1999)). First some terminology. A DRS consists of two compartments. The top compartment, $U(K)$, the so-called universe, contains the discourse referents. The bottom part, $Con(K)$ contains the conditions, which are either atomic formulas (e.g. $\texttt{see}(\texttt{y},\texttt{x})$), or complex ones containing subDRSs (e.g. $\neg K'$ or $K' \rightarrow K''$). An intensional model is a tuple $\langle D, W, I \rangle$. A central notion in DRT semantics is that of a verifying embedding, which is a partial function from the set of discourse referents to D. A DRS K is true in w relative to anchor f iff there is an extension of the anchor to $U(K)$ that verifies K in w. Notation:

(5) $[\![K]\!]^f_w = 1$ iff there is an embedding $g \supseteq f$ with $Dom(g) = U(K)$ and $g \models_w K$

An embedding g verifies K in w iff it verifies all conditions of K:

(6) $g \models_w K$ iff for all $\psi \in Con(K)$: $g \models_w \psi$

Condition verification, finally, is defined by cases. Here is an example of an atomic and a complex condition:

(7) a. $g \models_w P(x_1, \ldots x_n)$ iff $\langle g(x_1), \ldots, g(x_n) \rangle \in I_w(P)$
b. $g \models_w \neg K'$ iff there is no $h \supseteq g$ with $Dom(h) = Dom(g) \cup U(K')$ and $h \models_w K'$

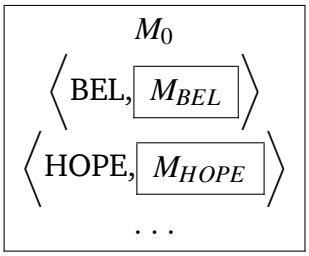
We can now define the important notion of a proposition expressed by a DRS relative to an anchor:

(8) $$[\![K]\!]^f = \left\{ w \in W \;\middle|\; [\![K]\!]^f_w = 1 \right\}$$

Now let's turn to the interpretation of attitudes in a mental state description. A mental state description contains descriptions of the various attitudes in the form of DRSs. I use the following notation, if M is a mental state description, M_{BEL} is the DRS that is paired with the label BEL within M;

M_{HOPE} the DRS labeled HOPE and so on. M_0 will denote the remainder, i.e. the global mental file cabinet.

(9) $M =$



We can define the wide belief or hope content of M relative to external anchor f as follows:

(10) a. $[\![M]\!]^f_{BEL} = [\![M_{BEL}]\!]^f$
b. $[\![M]\!]^f_{HOPE} = [\![M_{HOPE}]\!]^f$

The wide content of my hope in the letter example is then the set of worlds in which the actual letter is a notification of acceptance. Pierre's belief content is the set of worlds w such that the actual city, London, is pretty in w and not pretty in w, i.e. the empty set.

Psychologically speaking, whether or not M accurately describes someone's mental state has nothing to do with these singular propositions. As announced in section 2.3.2, to capture the psychological interpretation we need a different notion of attitude content, narrow content. In determining, say, narrow hope content, the free discourse referents in the belief box should get their reference fixed not by the external anchor, but by the descriptive content in the mental files.

Sticking in the anchoring metaphor we want to define the *internal anchor* determined by mental state M as an embedding from mental file discourse referents to entities that satisfies all the conditions in M_0. However, a set of conditions, as in M_0, is not satisfied by a mere sequence of individuals – we always need a possible world coordinate in order to evaluate DRS conditions. Or rather, to fix also the reference of the non-descriptive self-file i, a centered world c ($=\langle w_c, a_c\rangle$). The relativized definition of an internal anchor, determined by M, relative to c, becomes:

(11) $Anch(M,c)$ is the unique embedding $g : U(M_0) \to D$ with $g(\mathtt{i}) = a_c$ that verifies M_0 in w_c.

Note that $Anch(M,c)$ is undefined if there is no unique such embedding of M_0 in w_c, i.e. if M_0 doesn't determine a unique sequence of objects in w_c that

satisfies all descriptive mental file conditions in w_c.[7]

As discussed in 2.3.3, an agent presumes the contents of the mental file cabinet to correspond to – or at least include, cf. footnote 6 – the acquaintance relations between her and a number of *res*. Let's assume furthermore that she presumes these acquaintance relations to be descriptively rich enough to pick out these *res* uniquely.[8] Formally, this means that for any doxastic alternative c the descriptive content in M_0 must be rich enough to determine a unique verifying embedding relative to c. In other words, M_0 is a correct description of a's mental files in w iff for any $c \in Dox(a,w)$, the internal anchor $Anch(M,c)$ is defined.[9]

The idea of an internal anchor was so that we can compute the content of the attitude boxes. Take hope: M correctly represents the agent's hopes if (i) the agent believes her mental files to refer uniquely, i.e. for all doxastic alternatives, an internal anchor is defined, and (ii), given a doxastic alternative c, the agent's hope alternatives are compatible with the proposition expressed by the hope box relative to the internal anchor determined by c.

7 In line with the common intuition that mental files correspond to conceptual individuals, we can define from $Anch(M,c)$ the notion of an internal anchor as such, $Anch(M)$, which is a (partial) mapping from discourse referents to (partial) individual concepts (functions from centered worlds to individuals, $\in D^C$):

$$Anch(M) = \text{the } f : U(M_0) \to D^C \text{ s.t. for all } x \in U(M_0), c \in C\text{: } f(x)(c) = Anch(M,c)(x).$$

In this way, a mental state description M effectively associates with each mental file an individual concept. For instance, in the letter example, `x` is associated with the concept of a letter that the agent sees in her pigeon hole, and the self-file `i` is, always, associated with the self-concept, i.e., the function that maps any centered world c to its center coordinate a_c. Cf. Zeevat's (1999) closely related notion of intensional anchors, or Yanovich's (2011) notion of characters.

8 That is, with respect to an agent's doxastic alternatives, her acquaintance relations behave like functions (i.e., in any doxastic alternative, the agent can't stand in acquaintance relation R to two distinct objects at the same time). On this assumption, acquaintance relations correspond roughly to what Kaplan (1968) calls vivid names for an agent.

9 A reviewer points out a potential counterexample involving an individual who tracks several similar objects simultaneously, say, a bunch of moving yellow dots on a screen. At the DRS-level there could be multiple distinct files (`x`,`y`,...) with the same contents (`see(i,x)`,`yellow.dot(x)`,`see(i,y)`,`yellow.dot(y)`,...), externally anchored to distinct objects. But when we then try to determine $Anch(M,c)$ we run into trouble because there are multiple ways of associating both `x` and `y` with a "moving yellow dot that I see now". So, our semantics doesn't allow us to interpret mental state descriptions with multiple files with the same content. An obvious solution would be to assume that in fact the pieces of descriptive, reference fixing information associated with `x` and `y` may be similar but not really identical. For instance, the mental file cabinet in this case could plausibly contain the information that `x` is currently located to the left of `y`.

And similarly for the other attitudes:

(12) a. M correctly represents a's beliefs in w iff for all $c \in Dox(a,w)$: $Anch(M,c)$ is defined and $Dox(a,w) \subseteq [\![M_{BEL}]\!]^{Anch(M,c)}$

b. M correctly represents a's hopes in w iff for all $c \in Dox(a,w)$: $Anch(M,c)$ is defined and $Hope(a,w) \subseteq [\![M_{HOPE}]\!]^{Anch(M,c)}$

Thus, given an intensional model that specifies doxastic and other attitudes of agents as sets of centered worlds, we say that a mental state description M partially represents the complex mental state of agent a in w iff M correctly represents a's beliefs, hopes, desires, intentions, etc., as defined in (12).

3 Participant-neutral interpretation: Updating the common ground

With the file-based mental state descriptions in place, we now turn our attention to linguistic communication. My ultimate aim is to build a precise, formal semantic model of the traditional view of communication in which the speaker linguistically encodes a belief in the form of a sentence, so that the hearer who receives the sentence can decode it to get at the speaker's original belief. We now have a way to formally represent the beliefs and other attitudes of both speaker and hearer, viz., as parts of mental state descriptions. We now need a theory of linguistic encoding and decoding. I provide such a theory using, again, the formal framework of DRT. In this section, I first introduce the standard DRT model of communication as it is typically used in linguistics. I adapt it to an asymmetric speaker–hearer model in section 4.

As the name suggests, DRT is primarily a theory of discourse interpretation, where a discourse is a series of utterances constituting a conversation between a speaker and a hearer. Following ideas of Stalnaker (1970) the goal of the utterances in a discourse is to effect a growth of information in the common ground. DRT provides a formal language for representing the common ground and a description of how sentences effect information growth in common ground.

By way of illustration, let's say we've been discussing farmer John, so the existence of a farmer named *John* has become firmly established in the common ground. We represent the relevant part of this common ground in standard DRT as follows:

(13)

x
farmer(x) name(x,John)

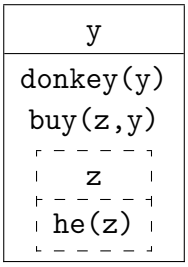
This well-formed formula of the DRS language represents the information that there exists an individual who is a farmer and who is named *John*. With the formal DRS syntax and model-theoretic semantics provided in section 2.4, we can make this precise: $[\![(13)]\!]$ = the set of possible worlds in which there is a farmer named John.

In this context a new sentence is uttered.

(14) He owns a donkey.

DRT aims to describe how this new sentence affects the common ground as represented in (13). In van der Sandt's (1992) presupposition-driven incarnation, context change is computed in two steps. The first step is to translate the sentence into a preliminary DRS, a logical representation of its context change potential. An important feature of the so-called construction algorithm is that it identifies a certain class of expressions as presupposition triggers. In (14) we see a third person pronoun *he*, which triggers[10] the presupposition that there exists a uniquely salient male, third person individual. Presuppositions are represented in the language of preliminary DRSs as free variables with presupposed content as conditions in a dashed box:

(15)



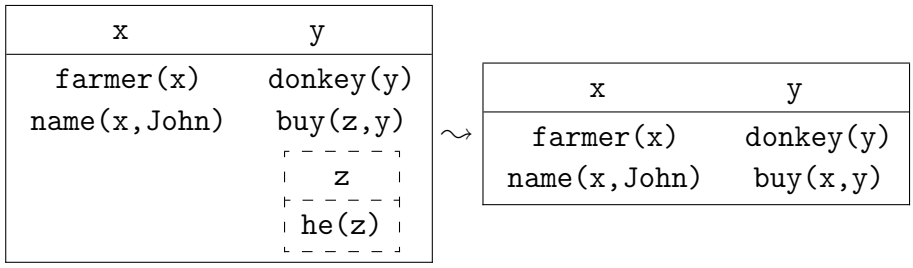
In words: there is a donkey, y, and z bought it, where z is a presupposed male third person individual. In other words, we treat the sentence in (14) as presupposing that there exist a male third person, while asserting that he bought a donkey.

The second step of the interpretation process is the resolution of the sentence's presuppositions by the resolution algorithm. We merge the context and the preliminary DRS (notation: (14)$\oplus$(15)) and then look for suitable antecedents for all presupposed discourse referents. In this case, the context provides a global discourse referent for a farmer John, which plausibly

10 A variety of linguistic tests can be used to establish whether a certain construction or lexical item is a presupposition trigger, and what presupposition it triggers. The classic test involves embedding under negation, i.e. *the King of France* presupposes the existence of a King of France because both *The King of France is bald* and *The King of France is not bald* imply that one exists. Cf. Geurts (1999) for a reliable "presupposition test battery".

matches the content (third person, male) of the presupposition triggered by the pronoun. Hence, we *bind* z to x:

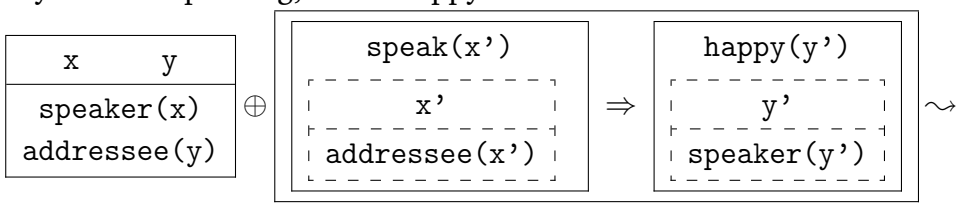
(16)

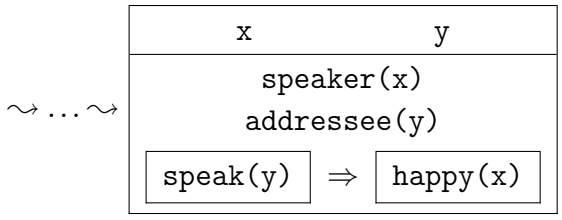


In general, presupposition resolution is a complicated process constrained by a variety of lexical, semantic and pragmatic factors. For details I refer the reader to Beaver & Geurts (2011) and references therein. For our current purposes, an important feature to note is the use of DRSs as representations of the common ground. Many linguistic phenomena in the semantics/pragmatics interface can be quite adequately captured from such a purely participant-neutral perspective. Even phenomena which are intricately related to perspective taking, such as the interpretation of indexicals and *de se* attitude reports, are typically analyzed in this way (cf. Zeevat 1999 and Maier 2010, respectively).

To bring out the difference between the participant-neutral and a speaker–hearer-oriented conception of communication, consider the interpretation of the indexicals *I* and *you* in the current model. The starting point is that *I* should be analyzed simply as triggering the presupposition that there exists a unique current speaker. If we assume that the speaker and hearer of any speech act are explicitly represented as salient individuals in the common ground prior to the interpretation of that speech act, this presupposition will always be globally bindable to the actual speaker. Consider an utterance of (17) in a minimal context, containing a salient speaker and hearer. After merging context and preliminary DRS the presuppositions triggered by *you* and *I* can be bound globally.

(17) If you were speaking, I'd be happy.





We thus derive a wide scope reading that seems to capture the correct truth conditions: there's an actual speaker and addressee, and if the latter were speaking, the former would be happy.[11] In sum, in the framework of DRT based on common ground updates, indexicals can be straightforwardly analyzed as presupposition triggers.

Despite the wide empirical coverage of this and other variants of dynamic semantics, Kamp and others point out that some phenomena can only be described properly by moving to an asymmetric model of communication that takes the differences between the speaker and hearer perspectives into account. A illustrative case in point – other than Stalnaker's puzzle about *de se* communication, which I take on in section 5 – are specific indefinites. Sæbø (2012) presents a scenario where he confesses to his wife, "I have met someone else". On the one hand, he is referring to a specific individual, one that he is so intimately acquainted with that his thoughts about her are singular, *de re* thoughts. What's more, Sæbøargues that the expression of this singular thought is likewise a singular proposition, by observing that in a report the indefinite can be replaced by a directly referential expression, as in "He told his wife that he has met me". Hence, the indefinite in the original confession must have been used as a referential expression. On the other hand, by choosing an indefinite rather than, say, her name, what his confession manages to convey to his wife is merely an existential proposition, viz. that there is someone else that he has met. The tension between specificity, or even direct reference, and existential quantification is not easily resolved in the common ground update conception of semantics – or, for that matter, in a static, proposition-based formalism. What we would want to say, according to Sæbø (2012), is that the indefinite *someone else* here is somehow directly referential "for the speaker", but at the same time merely existential "for the hearer".

The specific indefinites example above was meant to convince you that it

11 Kaplan (1989) argues that merely assigning wide scope is not enough to capture the interpretation of indexicals. For simple statements like *I am speaking*, the current proposal would indeed fail to account for our intuitions about the modal status of the proposition expressed. Cf. Maier (2009) and Hunter (2012) for some DRT extensions that bring genuine direct reference to (participant-neutral) DRT.

could be worthwhile for semantics to study communication asymmetrically, i.e. as transmission of information from speaker to hearer. In the remainder of this paper I will propose a way to make sense of such a speaker–hearer asymmetry with respect to communication.

4 Asymmetric semantics: distinguishing speaker and hearer

In this section we return to a traditional picture of communication alluded to in section 1. This involves describing linguistic meaning from two perspectives. There's the perspective of the speaker, who chooses a part of her mental state that she wants to communicate and tries to find the words to do so. And there's the perspective of the hearer, who receives an utterance and has to interpret it so he can update his own mental state accordingly.

In the following I discuss both perspectives within the general framework of DRT. To this end I combine the DRT-based formalism for describing mental states (section 2) with, for the hearer's perspective, the DRT-based dynamic presupposition theory (section 3).

4.1 The speaker's perspective

We describe linguistic communication from the speaker's perspective by defining a mapping from parts of mental state descriptions (as in section 2) to sentences – a sentence production algorithm. This is by far the most underdeveloped area within DRT research, and I will not contribute much here. What I will do is merely to discuss some specific examples so as to get a rough idea of what should go into such an algorithm.

Say, Pierre wants to express one of his beliefs about the wondrous city he read about as a child. Consider the mental state description of Pierre in (4) from section 2.3.2. What he wants to express is the proposition represented by the condition $\texttt{pretty(x)}$ in his belief box, where $\texttt{x}$ is the Londres-file, i.e. the mental file based on his acquaintance with London through reading a French children's book. A sentence production algorithm will, in some form or other, take into account the following factors. First, the fact that the relevant content is represented in the belief box will prompt the production of an indicative statement. Second, the atomic predicate–argument structure of $\texttt{pretty(x)}$ will trigger a subject–predicate sentence frame of the form $NP_{\texttt{x}}$ *is pretty*. Finally, the fact that $\texttt{x}$ is grounded in a mental file will trigger a search for an appropriate definite NP. What NP gets chosen depends first of all on the content of the file associated with $\texttt{x}$. In this case the file contains a name

predication name(x,Londres), which is enough to trigger the insertion of the mentioned proper name *Londres* in the NP slot: "Londres is pretty".

Another example. Say I want to express my hope about the letter as represented in (1) from section 2.2, i.e., acceptance(x). On the basis of its position within the mental state description, the production algorithm triggers a statement of the form *I hope that* NP_x *is a notification of acceptance*, where NP_x is some noun phrase that serves to pick out x, the letter. In this case the mental file in question does not contain a convenient name predication. Instead it describes the object as something in my immediate surroundings that I am currently looking at (look.at(i,x)). This looking arguably raises its salience in a way that tells the production algorithm to insert a proximal demonstrative *this*. Hence, I would utter: "I hope that this is a notification of acceptance."

The examples above illustrate the linguistic expression of a singular attitude. In both cases the choice of referential expression was guided solely by the contents of the speaker's mental files. In some cases however this choice is affected by pragmatic considerations, i.e. the speaker takes into account whether or not the chosen NP will have the desired effect in the hearer. For instance, Pierre's use of *Londres* may accurately reflect the contents of his mental file, but if he believes that his interlocutor does not have a mental file with a similar name predication, then he should probably refrain from using it. In such a case Pierre might choose a different description from his Londres file, one he does believe to share with his interlocutor, say *the city described in that book over there on the table is pretty*. If Pierre cannot find any shared common ground to pick out the specific city he has in mind, he could resort to a (specific) indefinite construction, as in *There's a city I used to read about as a kid. It was pretty.*. Modeling such pragmatic considerations about what the speaker believes about the addressee's beliefs goes well beyond the scope of this paper. The above serves merely to illustrate what kind of components should eventually go into a theory of the speaker's side of the communication of singular attitudes.

4.2 The hearer's perspective

The use of DRT to model interpretation of a discourse from the hearer's point of view, i.e. with context DRSs representing a hearer's mental state, is implicit in some early work on DRT. However, on closer examination the simple DRSs familiar from linguistics textbooks don't suffice as representations of a mental states. What I propose here is to combine the independently motivated mental state descriptions from section 2 with the dynamic, presupposition-driven

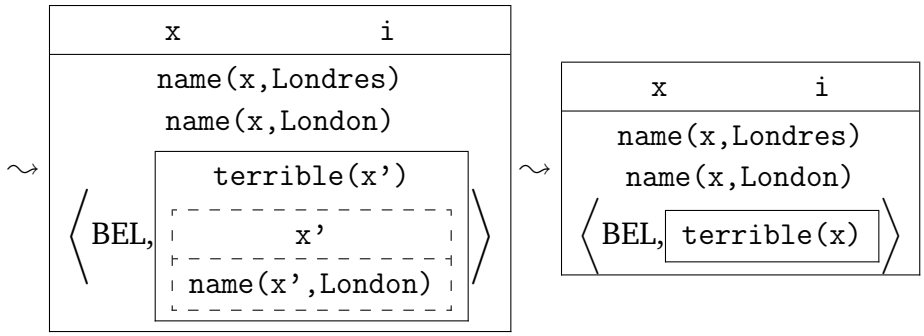
model of interpretation from section 3.

The key step is thus to replace simple DRSs with mental state descriptions as the contexts that get updated. The first stage of the interpretation process, i.e., the construction of a preliminary DRS from a sentence, can be imported as is. The difference lies in the second stage, starting already in the merge operation. Restricting ourselves to a cooperative information exchange, the rule will be that the hearer adds preliminary representations of utterances to his own stack of beliefs, i.e. as new conditions in her belief box. More precisely, (18) shows the first step in the interpretation process of a hearer with mental state M interpreting a preliminary DRS representation ψ of an utterance.

(18)
$$\boxed{\begin{array}{c} M_0 \\ \left\langle \text{BEL}, \boxed{M_{BEL}} \right\rangle \\ \left\langle \text{HOPE}, \boxed{M_{HOPE}} \right\rangle \\ \ldots \end{array}} \oplus \psi = \boxed{\begin{array}{c} M_0 \\ \left\langle \text{BEL}, \boxed{M_{BEL} \oplus \psi} \right\rangle \\ \left\langle \text{HOPE}, \boxed{M_{HOPE}} \right\rangle \\ \ldots \end{array}}$$

Let me illustrate this with Pierre saying to a French speaking friend "London est terrible." Pierre's interlocutor doesn't really know much about London so she trusts Pierre's judgment, but she does know that London and Londres refer to the same city. Her interpretation proceeds as follows: (i) construct a preliminary DRS, featuring a presuppositional representation of the proper name *London*, (ii) add it to the belief box, and (iii) resolve the name-presupposition by binding it to the London/Londres-file.

(19)
$$\boxed{\begin{array}{|cc|} \hline \texttt{x} & \texttt{i} \\ \hline \texttt{name(x,Londres)} & \\ \texttt{name(x,London)} & \\ \left\langle \text{BEL}, \boxed{} \right\rangle & \\ \hline \end{array}} \oplus \boxed{\begin{array}{|c|} \hline \texttt{terrible(x')} \\ \hline \texttt{x'} \\ \hline \texttt{name(x',London)} \\ \hline \end{array}}$$



By contrast, if the hearer only had a Londres-file, which does not contain the information that its referent is also known as *London* in English, she would be unable to bind the presupposition.[12] Finally, if the hearer were like Pierre in thinking that London and Londres are two different cities, there would be two mental files, and she would bind the presupposition to her London file.

In the next section I further illustrate the proposed communication model by applying it to *de se* beliefs and indexicals. I'll show how the puzzling asymmetry in communicating first person attitudes can be derived from the account sketched here, thus solving Stalnaker's puzzle from section 1.

5 *De se* communication revisited

We have developed a concrete model of the linguistic communication as the transmission of information from the speaker's to the hearer's mental states. In section 1 we saw how a simple account of meaning as propositions expressed runs into trouble when it comes to communicating *de se* beliefs. The current model is sufficiently expressive to deal with indexicals and *de se* attitudes. In this section I demonstrate how it effectively solves the problem of *de se* communication and how this solution relates to some previous proposals in the literature. I focus on the speaker's expression of first and second person *de se* beliefs, and the hearer's subsequent interpretation of utterances containing *I* and *you*.

12 In such a case she could resort to *accommodation*, i.e. she might trust Pierre to know what he's talking about and add the information that there is a city named *London* to her belief box. She might perhaps infer, based on phonological similarity and/or contextual clues, that this new name refers to the city she knows as Londres and thus equate the accommodated discourse referent with the discourse referent associated with her existing Londres file.

5.1 Communicating a first person thought

In the previous section I demonstrated how referential expressions get produced and interpreted. For the speaker, proper names, but also definite descriptions and even (specific) indefinites, are the verbalizations of mental files containing certain triggering conditions. For the hearer, all referential expressions are treated uniformly as presupposition triggers.

In this section we zoom in on *de se* attitudes and the production and interpretation of indexicals. We start with the communication of a first person belief via a first person pronoun.

In section 2.3.1 we represented Lingens's first person *de se* belief that he is a lost amnesiac. Let's say he wants to express his belief that he is lost, `lost(i)`, to the librarian. This case differs from those considered in section 4 in that the self-file `i` is non-descriptive. That is, one is acquainted with oneself in a direct way that does not involve a descriptive mode of presentation, so there need not be any descriptive conditions associated with `i`. As a result, in the case of a radical amnesiac like Lingens, there is no name condition or anything else that could constrain the choice of the subject term in $NP_{\texttt{i}}$ *be lost*. We therefore postulate a special *de se* production rule, mapping `i` to the first person pronoun *I* directly. This gives the expected production result: "I am lost".[13]

From a production perspective, *I* is a "*de se* pronoun", a way for the speaker to unambiguously express a *de se* attitude.[14] The crucial asymmetry, at the root of our solution to Stalnaker's puzzle, is that, as I will argue next, *I* does not have such a special first person *de se* status for the hearer.

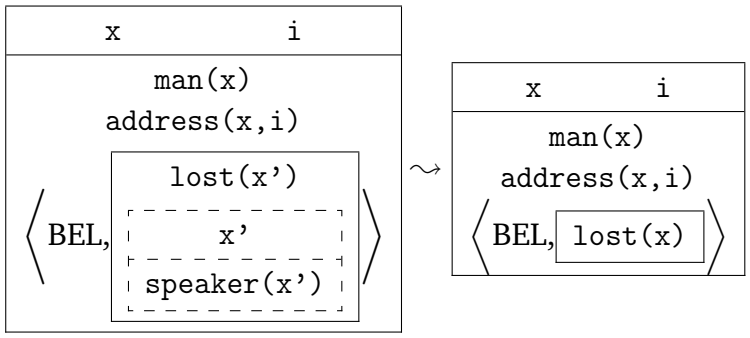
To illustrate the interpretation of indexicals, consider how the librarian would interpret Lingens's utterance of *I am lost*. Following the standard presuppositional theory of indexicals from section 3 we treat *I* as lexically trig-

13 I assume a similar rule in the temporal domain, mapping `n` to *now* and present tense morphology. Other indexicals are the result of spelling out descriptive files that involve these two pure mental indexicals in certain key conditions. We already saw *this* triggered by a file `x` containing `look.at(i,x)`. Similarly, *you* is inserted for a file `x` representing the center's addressee, i.e. with condition `address(i,x)`, and *here* is triggered by `located(i,x)` etc. The production rules for these impure indexicals crucially involve triggering conditions that express relations to `i` and/or `n`.

14 I'm scare quoting "*de se* pronoun" as that term is also used in a different sense, to describe elements like PRO or African logophors that force *de se* readings of reports (Schlenker 2003). Note that English *I* is not a *de se* pronoun in this traditional semantic sense. Reports in the first person, especially in the past tense, do allow non-*de se* readings in mistaken identity scenarios: *Listening to the election speeches on TV I thought that mine sounded great and I hoped that I would win, but I was so drunk that I didn't even recognize that I was looking at myself!*. Moreover, logophors and PRO are not *de se* pronouns in the newer sense.

gering the presupposition that there is a (uniquely salient, current) speaker. We'll assume that, prior to the interpretation of Lingens's utterance, the librarian's mental state already contains a file for a man standing in front of the desk, addressing him. Following the interpretation algorithm of section 4.2 we first add the preliminary DRS of the sentence to the librarian's belief box, and then we bind the speaker-presupposition to the file for the man addressing the subject:

(20)



Note that the pronoun *I* cannot be bound to the self-file `i`, as the interpreting agent is not currently a speaker.

Summing up: a speaker produces *I* to express *de se* attitudes involving the self-file `i`, while a hearer interprets *I* by constructing a lexically specified speaker-presupposition, and binding that to some mental file representation of the current speaker. Stalnaker's puzzling asymmetry in the communication of first person belief thus falls out naturally. In particular, the asymmetry derives from two independent theoretical assumptions of the current analysis of mental states and communication: (i) there's a direct, lexically encoded link between `i` and *I* in production, and (ii) all definite NPs, including indexicals, uniformly trigger descriptive presuppositions.

Observing and even formalizing a production–interpretation asymmetry with respect to the first person pronoun and *de se* attitudes is nothing new. Here is how (Kamp 1990: 69) puts it:

> There is an intimate connection between the meaning of "I" and the special access we have to ourselves, but this connection is restricted to the context of language production. For the interpreter the word "I" is much like a third person demonstrative such as "that man" or a deictic use of "him"

What is original about the current analysis of the first person is therefore not the special production rule for *I* – this is just the well-established doctrine of the essential indexical. Nor is it the fact that the hearer interprets *I* via a

descriptive representation of his addressee. Rather, what's new is the uniform mechanism by which this interpretation is derived. On the current proposal, the hearer's interpretation of *I* proceeds exactly like the interpretation of other definites, i.e. via the construction and resolution of a descriptive, existential presupposition.[15] Both the lexical content of the presupposition associated with *I* and the pragmatic/semantic resolution mechanism itself are independently motivated within the participant-neutral tradition of DRT (Zeevat 1999, Maier 2009, Hunter 2012).[16]

In the next subsection I will bring out a more substantial point of departure from competing asymmetric accounts of *de se* communication by examining the way my proposal extends from first to second person, and to cases where the interpreter is not the intended addressee.

5.2 Notes on the second person

Related proposals that assume a production–interpretation asymmetry of *I* often claim that *you* is somehow the mirror image of *I*. Consider for instance the continuation of the passage quoted above from (Kamp 1990: 69–70):

> [...] With 'you' the story is much the same, only reversed. 'You' also bears a special relationship to *i*, but here it is the construction rule, and not the verbalization rule that must exploit the special relation to the self. [...] [The construction rule for 'you'] can be succinctly stated as:
>
> Represent the referent of 'you' as *i*.

Or consider the following passage from Wechsler (2010), who proposes a non-DRT-based semantic account of asymmetric *de se* communication:

> Most work on self-ascription has focused on the first person, but second person pronouns have exactly the same self-ascriptive force, only applied to the addressee instead of the speaker.

15 The uniform presupposition-driven analysis of the hearer's perspective distinguishes the current proposal from a recent alternative solution to Stalnaker's puzzle: Weber's (2012) "recentering" approach. Like the current proposal, Weber's relies on distinguishing the speaker's and hearer's perspectives in communication. However, where I rely on the general mechanisms of presupposition resolution, his model of the hearer's interpretation involves a (arguably more ad hoc) mechanism dedicated to "recentering" the content expressed by the speaker.

16 For a related, but not DRT-based, presuppositional analysis of indexicals see García-Carpintero (2000).

In other words, according to these authors there is a special lexical rule for the interpretation of *you* that mirrors the special production rule for *I*. In our terminology, this special interpretation rule would directly map *you* to `i` in the interpretation process. I will show below that no such special treatment of *you* is needed, nor, in fact, desirable.

First, let me illustrate my proposal with a second person continuation of the Lingens example. Having just learned that the man in front of him is lost, the librarian looks through his mental file cabinet for location information. Let's assume that the librarian finds a file `y` for Stanford with the information that he (`i`) and his interlocutor (`x`) are located there. The production algorithm turns the relevant condition `in(x,y)` into something like $NP_{\texttt{x}}$ *be in* $NP_{\texttt{y}}$. Like names and definite descriptions, the second person pronoun, *you*, is lexically triggered by a descriptive condition occurring in a mental file – in this case it's `address(i,x)`.[17] Given the mental state description below, the production algorithm will then insert *you* for $NP_{\texttt{x}}$, and, based on `name(x,Stanford)`, it inserts *Stanford* for $NP_{\texttt{y}}$.

(21)

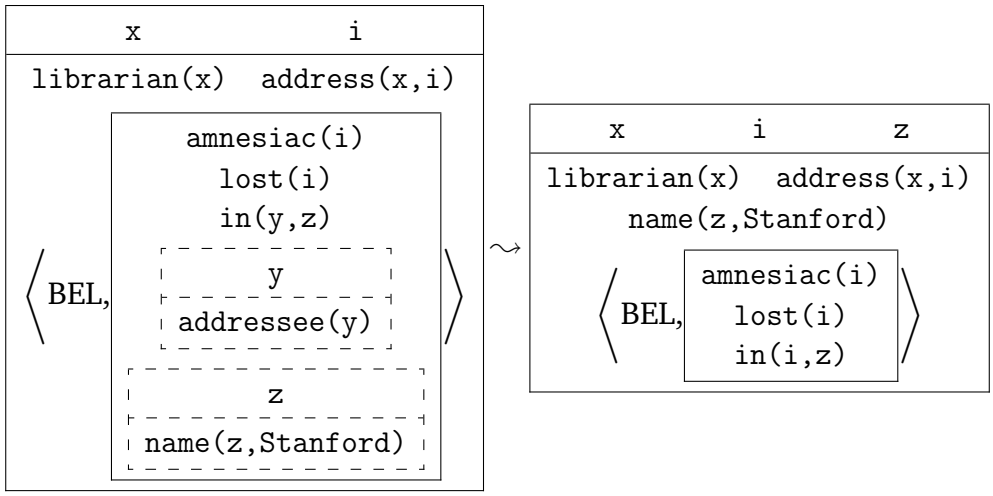
x	i	y
man(x)	name(y,Stanford)	
address(i,x)	in(x,y)	
⟨BEL, lost(x)⟩		

$\rightsquigarrow$ "You are in Stanford"

Now for the interpretation side. Lingens hears the librarian say "You are in Stanford". He computes the preliminary DRS, treating both *you* and *Stanford* as presupposition triggers, and adds that to his belief box. The addressee-presupposition binds to `i` since `i` occurs as the second argument to `address`. The name presupposition binds to Lingens's file on that city (or else, such a named file would be accommodated).

17 `address(i,x)` $\approx$ "I am going to address x with the speech act under construction." A proper formalization, unifying the interpretation of this predicate for speaker and hearer would require that we introduce a third argument that explicitly links it to (a file representing) the utterance currently being interpreted or produced.

(22)

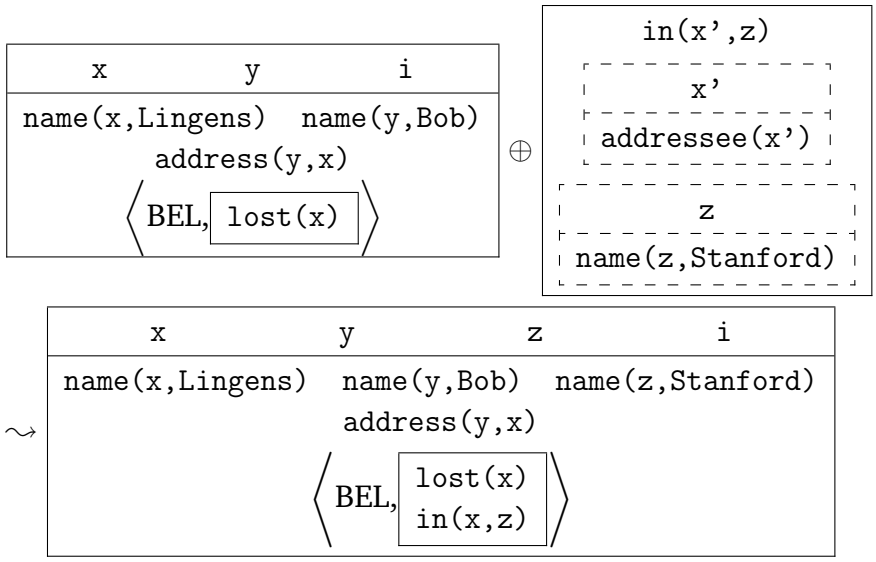


We see here that an utterance of *you* does eventually get associated with the hearer's self-file, `i`, just as Kamp and Wechsler postulated. However, unlike with the production of *I*, this link between *you* and `i` is not directly stipulated in the lexicon, but rather the result of the usual presupposition triggering and semantic/pragmatic resolution process. In sum, both the production and the interpretation of *you* proceed like that of other referential expressions.

For the example above we derived the same eventual output as Wechsler or Kamp, but in some situations our predictions diverge. As I will demonstrate next, only the current proposal automatically derives the right readings for cases of eavesdropping and some cases of miscommunication in which either the addressee fails to realize she is the addressee, or someone other than the addressee falsely believes to be the addressee.

Let's start by examining our predictions for an eavesdropping scenario. NSA agent Mary has hacked Lingens's phone and is able to remotely overhear the conversation between Lingens and the librarian, Bob. She knows both men by name and knows that Bob is addressing Lingens when he says "You are in Stanford". Her interpretation of this utterance proceeds exactly like Lingens's, viz. by constructing the exact same preliminary DRS and then updating her mental state with it.

(23)



That is, Mary interprets Bob's words as inducing in her the belief that Lingens is in Stanford.

By contrast, the simple rule *represent the referent of 'you' as* `i` from the Kamp quote above clearly makes the wrong prediction. Mary does not conclude from Bob's utterance that she herself is in Stanford. At the very least Kamp's rule would have to be restricted to the addressee. Completely new rules would then have to be stipulated to describe the interpretation of *you* by third parties.[18] Considering the two types of miscommunication alluded to above would further complicate this extension.

Consider the case of a third person who falsely believes she is the addressee. Just before Lingens came up to him, Bob was talking to his girlfriend on the phone, giving her directions. In fact, he didn't end this call before answering Lingens. So when his girlfriend hears him say "You are in Stanford", she thinks he is still talking to her. In this situation, the girlfriend has a mental file `x` for her boyfriend Bob with a condition $\texttt{address}(\texttt{x},\texttt{i})$. Hence, when she updates her mental state description with the preliminary DRS for the sentence, the addressee-presupposition will naturally bind to `i`. In this way we correctly predict that, as a result of interpreting Bob's utterance, she comes to self-ascribe being in Stanford.

Finally, consider the inverse miscommunication: someone failing to realize he's being addressed. Lingens overheard Bob talking to his girlfriend and falsely assumes that he is still talking to her when he says "You are in Stanford". In this case, Lingens, who has formed a mental file for Bob's girlfriend,

18 Kamp (1990:69-70) is aware of this limitation and admits that other rules should be added.

representing her as *the person the librarian is talking to over the phone*, adds the same preliminary DRS as before. The addressee-presupposition will bind to the girlfriend file, because she is the most salient person currently being addressed – in Lingens's mental state description. We correctly predict that Lingens interprets Bob's utterance as meaning that the person on the other end of the phone is in Stanford.

In both cases, the presuppositional proposal makes the right prediction out of the box. The presupposition resolution algorithm will find the right antecedent in the mental file cabinet of the hearer, whether this hearer is being addressed or not, and whether she knows this or not.

By contrast, to save the Kamp/Wechsler proposal for *you*, we would have to further complicate it by postulating that the simple *de se* rule applies only to individuals who *think* they are being addressed, while a different rule applies to individuals who *think* they are merely overhearing the utterance they are interpreting.

6 Conclusion

In this paper I have presented two distinct applications of the general logical framework of DRT.

The first is a theory of the representation of complex mental states. It analyzes singular attitudes via descriptive mental files that are externally anchored to objects in the world. A special, non-descriptive file `i` represents the *de se* center of the subject's beliefs. Different attitudes are represented as distinctly labeled DRSs that all have access to shared discourse referents representing the mental files. A possible worlds interpretation maps these complex mental state descriptions to properties that an agent self-ascribes. The resulting framework is expressive enough to describe doxastic and non-doxastic attitudes, *de se* attitudes, double vision situations and even some cases of faulty perception and hallucination.

The second application of DRT is as a theory of discourse semantics. In linguistic applications, a DRS is used to represent the information that is common ground between speaker and hearer at some point in a discourse. Interpretation is then the process by which an utterance adds information to the common ground. DRT's formalization of this process involves the compositional construction of a highly underspecified preliminary DRS and a resolution algorithm that integrates the preliminary DRS with the context DRS by binding or accommodating presuppositions.

I bring these two distinct applications of DRT together in an account of linguistic communication that clearly separates the speaker's production of

an utterance from the hearer's interpretation. Such an asymmetric account consists in providing a production algorithm, mapping the speaker's mental state to a sentence, and an interpretation algorithm, mapping sentences to "belief change potentials". In my proposal, the mental states of speaker and hearer are modeled in an extension of DRT with mental files and attitudes, and the hearer's change in belief is modeled in terms of presupposition resolution.

On the speaker's side, an important component of the production algorithm is the mapping from mental files to referential expressions. The speaker's choice of referential expression is guided by, among other things, the presence of certain descriptive predicates in the file. For instance, a predicate `name(x,Mary)` triggers the choice of a proper name *Mary* to represent mental file `x`. The non-descriptive self-file `i` receives special treatment and gets mapped directly onto the first person pronoun *I*.

On the hearer's side, I adopt the theory of presupposition resolution in DRT. The only difference is that, in the current setting, a preliminary DRS is not meant to update a representation of the common ground, but the belief compartment in a representation of the hearer's mental state. An important feature of the resulting account is that all referential expressions, from definite descriptions to indexicals, are analyzed uniformly as presupposition triggers.

This theory of communication solves an old problem regarding the communication of *de se* attitudes: How come that when I communicate a first person *de se* belief, with a first person pronoun, my addressee will form a different, second person belief? More succinctly put, why is my *I* your *you*? The general account of communication provided here shows that this asymmetry in the production and interpretation of the first person derives from two independently motivated assumptions: for the speaker, *I* is directly linked to the self-file, but for the hearer, *I* triggers a descriptive presupposition that binds to the mental file representing the most salient current speaker.

References

Asher, Nicholas. 1986. Belief in Discourse Representation Theory. *Journal of Philosophical Logic* 15(2). 127–189.

Beaver, David & Bart Geurts. 2011. Presupposition. *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/presupposition/.

Evans, Gareth. 1982. *Varieties of Reference*. Oxford: Oxford University Press.

Fauconnier, Gilles. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.

Frege, Gottlob. 1918. Der Gedanke. *Beiträge zur Philosophie des deutschen Idealismus* 1. 8–77.
García-Carpintero, Manuel. 2000. A presuppositional account of reference fixing. *Journal of Philosophy* 97(3). 109–147.
García-Carpintero, Manuel. 2010. Fictional singular imaginings. In Robin Jeshion (ed.), *New essays on singular thought*, 273—-299. Oxford: Oxford University Press.
Geurts, Bart. 1999. *Presuppositions and Pronouns*. Amsterdam: Elsevier.
Grice, Herbert Paul. 1969. Vacuous Names. In Donald Davidson & Jaakko Hintikka (eds.), *Words and Objections*, 118–145. Dordrecht: Reidel.
Hintikka, Jaakko. 1969. Semantics for propositional attitudes. In *Models for Modalities*, 87–112. Dordrecht: Reidel.
Hunter, Julie. 2012. Presuppositional Indexicals. *Journal of Semantics* 30(3). 381–421. http://dx.doi.org/10.1093/jos/ffs013.
Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Centre Tracts.
Kamp, Hans. 1990. Prolegomena to a Structural Account of Belief and Other Attitudes. In Anthony Anderson & Joseph Owens (eds.), *Propositional attitudes: The role of content in logic, language, and mind*, 27–90. Stanford: CSLI.
Kamp, Hans. 2011. Representing De Se Thoughts and their Reports. http://nasslli2012.com/files/kamp_2011.pdf.
Kamp, Hans, Josef van Genabith & Uwe Reyle. 2003. Discourse Representation Theory. In Dov Gabbay & Franz Guenthner (eds.), *Handbook of philosophical logic*, vol. 10, 125–394. Heidelberg: Springer. http://dx.doi.org/10.1007/978-94-007-0485-5_3.
Kamp, Hans & Uwe Reyle. 1993. *From Discourse to Logic: an Introduction to Modeltheoretic Semantics in Natural Language, Formal Logic and Discourse Representation Theory*, vol. 1. Dordrecht: Kluwer.
Kaplan, David. 1968. Quantifying in. *Synthese* 19(1-2). 178–214. http://dx.doi.org/10.1007/BF00568057.
Kaplan, David. 1989. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 481–614. New York: Oxford University Press.
Kripke, Saul. 1979. A Puzzle about Belief. In A. Margalit (ed.), *Meaning and Use*, 239–283. Dordrecht: Reidel.
Lewis, David. 1979. Attitudes de dicto and de se. *The Philosophical Review* 88(4). 513–543. http://www.jstor.org/stable/2184843.
Maier, Emar. 2009. Proper Names and Indexicals Trigger Rigid Presupposi-

tions. *Journal of Semantics* 26(3). 253–315. http://dx.doi.org/10.1093/jos/ffp006.
Maier, Emar. 2010. Presupposing acquaintance: a unified semantics for de dicto, de re and de se belief reports. *Linguistics and Philosophy* 32(5). 429–474. http://dx.doi.org/10.1007/s10988-010-9065-2.
Montague, Richard. 1973. The Proper Treatment of Quantification in Ordinary English. In *Approaches to Natural Language*, vol. 49, 221–242. Dordrecht: Reidel.
Ninan, Dilip. 2010. De Se Attitudes: Ascription and Communication. *Philosophy Compass* 5(7). 551–567. http://dx.doi.org/10.1111/j.1747-9991.2010.00290.x.
Ninan, Dilip. 2014. On Recanati's *Mental Files*. *Inquiry* 1–9.
Perry, John. 1980. A Problem about Continued Belief. *Pacific Philosophical Quarterly* 61. 317–332.
Perry, John. 2003. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT press.
Pryor, Jim. 2013. Acquaintance, mental files and mental graphs. http://www.jimpryor.net/research/papers/Graphs.txt.
Recanati, François. 2012. *Mental files*. Oxford: Oxford University Press.
Sæbø, Kjell Johan. 2012. Reports of Specific Indefinites. *Journal of Semantics* 30(3). 267–314. http://dx.doi.org/10.1093/jos/ffs015.
van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377. http://dx.doi.org/10.1093/jos/9.4.333.
Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26(1). 29–120. http://dx.doi.org/10.1023/A:1022225203544.
Stalnaker, Robert. 1970. Pragmatics. *Synthese* 22(1-2). 272–289.
Stalnaker, Robert. 1981. Indexical belief. *Synthese* 49. 129–151.
Weber, Clas. 2012. Centered communication. *Philosophical Studies* 166(S1). 205–223. http://dx.doi.org/10.1007/s11098-012-0066-6.
Wechsler, S. 2010. What 'you' and 'I' mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86(2). 332–365. http://dx.doi.org/10.1353/lan.0.0220.
Yanovich, Igor. 2011. The problem of counterfactual de re attitudes. *Semantics and Linguistic Theory (SALT)* 21. 56–75. http://elanguage.net/journals/index.php/salt/article/view/21.56.
Zeevat, Henk. 1999. Demonstratives in Discourse. *Journal of Semantics* 16(4). 279–313. http://dx.doi.org/10.1093/jos/16.4.279.

Emar Maier
University of Groningen
Department of Philosophy
Oude Boteringestraat 52
9712GL Groningen
The Netherlands
emar.maier@gmail.com
sites.google.com/site/emarmaier