<http://www.diva-portal.org>

Postprint

This is the accepted version of a chapter published in *About oneself: De Se Thought and Communication*.

Citation for the original published chapter:

Pagin, P. (2016)

De Se Communication: Centered or Uncentered?.

In: Manuel García-Carpintero, Stephan Torre (ed.), *About oneself: De Se Thought and Communication* (pp. 272-306). Oxford: Oxford University Press

<https://doi.org/10.1093/acprof:oso/9780198713265.003.0011>

N.B. When citing this work, cite the original published chapter.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-138274>

De se communication: Centered or Uncentered?*

Peter Pagin

August 2, 2015

Abstract

It was pointed out, first by Robert Stalnaker, then also by Andy Egan, that David Lewis's model of *centered-worlds contents* has undesired consequences for communication of *de se* contents. The recent years have seen a number of attempts to save the model by amending it to handle *de se* communication. Proposals include the appeal to sequences of individuals in the centers, to *ersatz* classical propositions, and to operations of "re-centering". The authors are Dilip Ninan and Stephan Torre (sequences), Sarah Moss and Max Kölbel (*ersatz*), and Alan Gibbard and Clas Weber (re-centering). The present paper discusses these attempts. The conclusion is that they fail.

Keywords: Centered worlds; communication; content; David Lewis; *de se*; Robert Stalnaker.

1. Introduction

Frege held that first-person thoughts cannot be communicated:

Now everyone is presented to himself in a special and primitive way, in which he is presented to no one else. So, when Dr Lauben has the thought that he was wounded, he will probably be basing it on this primitive way in which he is presented to himself. And only Dr Lauben himself can grasp thoughts specified in this way (Frege 1918, p. 333).

*This material derives from a talk given at the Barcelona conference on *de se* communication, organized by Manuel García-Carpintero and Stephan Torre, in June 2012. The paper version, which in some respects differs substantially from the talk content, was discussed at a seminar in Barcelona in June 2014. I am grateful for comments on both occasions, to Francois Recanati, Dilip Ninan, Stephan Torre, Carla Merino Rajme, Josep Macià, Max Kölbel, and Dan Zeman. I am especially indebted to comments from Manuel García-Carpintero on the first draft of this paper, to both editors and an anonymous referee for comments on the second, and for discussion to Kathrin Glüer and Francois Recanati. This research has been funded by an ESF Eurocores project grant for the CCCOM project (*Communication in Context: Shared Understanding in a Complex World*), led by Åsa Wikforss.

This raises the question of how speaker can convey thoughts about himself to others. Frege proposed a solution. The passage continues:

But now he may want to communicate with others. He cannot communicate a thought he alone can grasp. Therefore, if he now says 'I was wounded', he must use 'I' in a sense which can be grasped by others, perhaps in the sense of 'he who is speaking to you at this moment'; by doing this he makes the conditions accompanying his utterance serve towards the expression of a thought (Frege 1918, p. 333).

Frege's solution is to appeal to an *ersatz* content: instead of thought content that only he himself can grasp, the speaker conveys a different but related content. The new content contains a mode of presentation different from that in the original content, but the object presented is the same: Dr Lauben. Lauben achieves this by attaching a different sense to *I*. I shall refer to the solution of finding a related *non-first-person* content as the *ersatz solution*.

John Perry (1979) rejected not only Frege's solution but also the premise that motivated it, that there are contents available only to privileged subjects. Perry (1979, pp. 15-6) called this view "limited accessibility". In his paper, Perry discusses examples that indeed can motivate a Fregean view of limited accessibility, but proposes a different way of handling them.

The examples illustrate what John Perry (1979, p. 5) called "locating" belief and which later have come to be known as "self-locating" beliefs. In the first of these examples (1979, p. 3), John Perry himself is in a supermarket, following a trail of sugar, trying to catch up with the shopper with a torn sack, the shopper who is making a mess. Eventually, after failing to catch up, Perry realizes that this shopper is he himself. Only after realizing this does he take the action of rearranging the sack in his cart.

Discussing the case, Perry asks what proposition he expresses by means of

(1) I am making a mess.

considered as uttered by Perry in the supermarket. He argues that there is no concept α such that α is uniquely true of Perry and such that the proposition Perry can express by (1) is *that α is making a mess*. The reason is that whatever α is Perry might fail to believe that *he is α* (1979, p. 7). Lacking that belief, the proper action of rearranging the sack will not result.¹

But this seems to lead back to the Fregean view of limited accessibility, for there seems then not to be any way for anyone else than Perry to believe exactly what he can

¹Perry (1979, pp. 9-12) also considers *de re* and demonstrative beliefs, and argues convincingly that these do not help us to get rid of the problem.

express by means of (1) at the time. As a way out, Perry considers the idea of *relativized propositions*, in particular propositions that are relativized not only to a possible world, but also to a person and a time (1979, pp. 13-5). Perry argues that this does not solve the problem either, for believing such a proposition is believing it true *for an individual at a time*. Since others can think the same relativized proposition as true for Perry-the-shopper-at-the-time, we still lack an explanation why Perry's belief leads to rearranging the sack but believing it by others don't have that effect. To rule this possibility out, we would need to treat the relativized proposition available only to Perry-the-shopper. But that again, Perry argues, leaves us with limited accessibility.

This argument against relativized propositions is flawed, however, as is evidenced by David Lewis's (1979) proposal of relativized propositions to deal with self-locating beliefs, or beliefs *de se*. After a suggestion by Quine (1969), Lewis proposes thought contents as sets of *centered possible worlds*. A centered possible world is a pair $\langle C, W \rangle$ of a center C and an ordinary world w (Lewis 1979, pp. 531-2). A center C in turn is either a space-time point, i.e. a pair $\langle l, t \rangle$ of a location and a time, or else, as is more common in current discussions, a pair $\langle I, t \rangle$ of an individual and a time.

Terminologically, in (1979) Lewis distinguishes between propositions, conceived of as sets of ordinary worlds, and sets of centered worlds, which correspond to properties, but the contemporary discussion regards the idea of a set of centered worlds as an alternative conception of what a proposition is.² Either way, a belief *de se* for Lewis (1979, p. 539) is the *self-ascription* of a property, which is a set of centered worlds. If, and only if, Lewis believes that *he* is making a mess at time t , he self-ascribes the set of centered worlds $\langle \langle I, t \rangle, w \rangle$ such that I is making a mess at t in w . Thus, the connection between relativized proposition, the set of centered worlds, and the believer, is the believer's act of self-ascribing the set. On this view, ascribing any property to anyone else proceeds via ascribing a property to oneself.

This is why Perry's argument fails. Only Perry can self-ascribe the set of centered worlds to *Perry* where the center is making a mess, but the *content* is equally accessible to everyone. Lewis can self-ascribe the same set of centered worlds to *Lewis*. There is therefore no question of limited accessibility. With the idea of taking a relativized proposition to be *true for I at t*, Perry did not clearly separate the content, which is the relativized proposition itself, and the belief-relation of the content. Rather, he in effect treated the relativized proposition as a propositional function that, when applied to an individual and a time, yields an ordinary possible-worlds proposition, which is then subject to ordinary belief. Lewis, with his idea of self-ascription as the primitive relation between a

²As pointed out by the editors, Lewis in later work (e.g. Lewis 2004) he also speaks of *egocentric propositions*.

believer and a content, saw it more clearly.³

However, this very idea creates a problem in connection with communication. The problem was first pointed out by (Stalnaker (1981, pp. 145-7)), and has more recently been emphasized by Andy Egan (2007). The problem is that because believing contents amounts to self-ascription of the contents, for a hearer to believe the content that a speaker has asserted amounts to *self-ascribing* the content, not ascribing it to the speaker. When Perry-the-shopper meets Lewis-the-shopper in the supermarket and utters (1), for Lewis to believe Perry is to accept the content Perry expresses, but accepting it amounts to self-ascribing it. That means that Lewis, in accepting Perry's assertion, ascribes to *himself* the property of making a mess.

This is of course unacceptable. So there is a problem. But what is the problem exactly? In the first place, we have the problem that we cannot combine the theory of centered-worlds propositions and self-ascription with the traditional idea of communication as transmission of content, i.e. with the idea that communication succeeds iff the hearer's resulting content is the same as, or sufficiently similar to, the speaker's original content. It may be said here that this is not Lewis's problem, since he of course did not, or would not, accept the transmission model of communication. I'll quote here from the report on an earlier version of this paper by an anonymous reviewer:

Semantics gives us truth conditions for statements, relative to a context. I use the semantics to figure out what my context must be like in order for your statement to be true. Given the conventions of truth and trust, I then come to believe that I am in a context of that kind. So if Perry say "I am making a mess" Lewis infers from the semantics that what he says is true if and only if Perry is making a mess, and so comes to self-ascribe the property of being in a world in which Perry is making a mess. So Lewis explicitly rejects the transmission model (Anonymous report, p 1).

I have no objection to what is said in this passage, only to the *non-sequitur* that once you have given up the transmission model, the problem is thereby eliminated. What remains to answer is the question: *How does it work?* In the first place, this is a *challenge* to provide an account of the new combination of semantics, centered-worlds propositions, self-ascription, and communication. In the second place, there is a *problem* if this challenge cannot be easily met and it remains unclear that it can be met at all. In this paper, I argue that the appeal to centered-worlds proposition does have *this* (latter) problem.

Friends of centered-worlds propositions have taken up the challenge. They have proposed amendments to Lewis's theory to get around the initial communication problem.

³A similar characterization of Perry's argument is given García-Carpintero 2014 (forthcoming).

They have done so in three different ways: some (Ninan, Torre) have proposed modifications of the theory of centered-worlds propositions themselves. Some (Moss, Kölbel) have proposed that the speaker thinks a centered-worlds proposition but *communicates* a related *non-centered* proposition. Some again (Weber, Gibbard) have proposed that the hearer performs a *transformation* of the asserted centered content which makes it apt for the hearer to self-ascribe.

This paper is concerned with those proposals. In section 3, I discuss the appeal to conversational sequences that have been proposed by Dilip Ninan (2010) and Stephan Torre (2010). In section 4, I discuss the appeal to associated un-centered contents, *ersatz* contents, proposed by Sarah Moss (2012) and Max Kölbel (2013). In section 5, I discuss an idea proposed by Allan Gibbard (2012) and by Clas Weber (2013), according to which the hearer in accepting the content *transforms* it by means of an operation Weber calls “recentering”. In all of these cases, I argue that the amended theory, to the extent it does not reduce to a theory of non-centered contents, is flawed.

There is no common way in which they all are flawed. Rather, each has its own problems which depends on the particular modifications made. However, there is a deeper problem with Lewis’s theory, which comes to light in considering communication: the connection between the content and the thinker is not representational. That is, the speaker is not represented, neither explicitly nor implicitly, in the content of the thought. Rather, it is the *belief-mode*, or *belief-force*, in the form of self-ascription, that makes the connection. This means that a subject cannot merely *entertain* the (centered-worlds propositional) thought that she has a certain property; what she entertains is just the content of the representation of the *property*. This does not matter for *belief*, which by definition *has* force. But it does matter for communication, because the hearer does not automatically *believe* the speaker: typically we first *understand*, then perhaps form a new belief. This picture of communication is hard to combine with the theory of centered-worlds propositions.

With respect to this feature of Lewis’s theory, we can characterize the three kinds of amendments as follows: The first kind (conversational sequences) attempts to neutralize the effect of individual self-ascription by *enriching* the contents so that the speaker’s self-ascription is true iff the hearer’s is. The second kind (*ersatz* contents) attempts to get around the problem by appealing to a relation to contents that don’t have this shortcoming. And the third kind (recentering) simply *encounters* this problem, since the transformation made by the hearer presupposes that she believes something: understanding is not independent of belief.

I shall start, in the next section, by elaborating a bit on these general features of the

relation between content and force, which are crucial in connection with the idea of centered contents.

2. Content and force

Frege regarded judging as *passing* from the Thought to the truth value. He wrote:

A judgment for me is not the mere comprehension of a thought, but the acknowledgment of its truth (Frege 1892, 34, note 7).

Judging can be regarded as advancing from a thought to its truth value. Naturally, this cannot be a definition. Judging is something quite of its own kind and incomparable (Frege 1892, 35).

Since for Frege, the truth value is the Reference (*Bedeutung*) of a sentence, a judgment is an advance from Sense to Reference. In case the subject makes a mistake, it is not the actual Reference, but rather the Reference the subject *takes it* to have.

We can transform Frege's view slightly, in the following way: in judging that *p*, the subject advances from content that *p* to the relevant *point of evaluation*, with respect to which this content is either true or false. Frege countenanced exactly one point of evaluation: the World. In a sense, for Frege, judging is therefore *applying* a content to the World. It is not an application in the same sense as that of an application of a predicate to a singular term, which yields a sentence, or propositional function to a singular concept, which yields a proposition. But, in line with how Frege thought, it is analogous to the application of a concept to an object, which delivers a truth value, in case there is only one point of evaluation.

However, if belief content is thought of as a possible-worlds proposition, then each possible world is a possible point of evaluation. The *relevant* point of evaluation for any actual judgment, or assertion, is then the *actual world*. Believing that *p* is taking *p* to be actually true. You apply the content to the actual world. This is also Francois Recanati's view:

[...] a content is not enough; we need to connect that content with the actual world, via the assertive force of the utterance, in virtue of which the content is presented as characterizing that world (2007, p. 37).

If contents are *temporal* propositions, true or false with respect to world-time pairs, then all actual judgments are applications to the ordered pair of the actual world and a relevant time, usually the time at which the judgment is made. This is the point with respect to which a sentence, used in a context of utterance, has its truth value. David Kaplan (1989, p. 522) says

If c is a context, then an occurrence of ϕ in c is true iff the content expressed by ϕ in this context is true when evaluated with respect to the circumstance of the context (Kaplan 1989, 522).

This fits well with the idea that the judgment, or the assertoric force, provides the world and the time. Believing that p is taking p to *actually and currently true*. You apply the content to the actual-current world-time pair.

In perfect analogy, if contents are Lewisian centered-worlds propositions, true or false with respect to individual-time-world triples, then all actual judgments are applications to the triple of the subject, the current time, and the actual world. And this is basically Lewis's idea of self-ascription. Believing the centered-worlds proposition *that I am making a mess* is applying that proposition to the egocentric-current-actual triple of the subject, the current time and the actual world, which is equivalent to ascribing to oneself the property of making a mess at the current time in the actual world.

The relation is general: if the content of a judgment is a function from points of evaluation (indices) of some type to truth values, then a judgment is the very step of applying that content to the relevantly actual index. To repeat, it is important to separate the idea of *evaluating* a content at a point, and *enriching* the content by means of a representation of that point. We need to distinguish between applying a content to a the current time and enriching the content with a representation of the current time. In Kaplan's framework this is the difference between the contents expressed by

- (2) a. John is making a mess.
- b. John is making a mess now.

By means of (2a), according to Kaplan, the speaker asserts a temporal proposition, true or false at the current time and the actual world, and by means of (2b) the speaker asserts an eternal proposition that is true or false at the actual world. The contents are different, but the first, expressed by (2a) as uttered at the current time, is true at the current time in the actual world just if the second, expressed by (2b) as uttered at the current time, is true at the actual world (at any time).

Observing this distinction between enrichment and evaluation is crucial for understanding the idea of centered content.

3. Relativity to sequences

One idea that has been employed by several authors, originally suggested by Lewis himself (1983a, p. 398), is to enrich the *centers* of centered worlds to contain a *sequence* of individuals rather than an individual. In modelling communication between two per-

sons, what is needed then is a pair of individuals. Instead of a center consisting of an individual-time pair $\langle x, t \rangle$, we now have a triple $\langle x, y, t \rangle$ of an individual, an individual, and time. The basic idea is that we fix an order between the two participants X and Y of the conversation, and then interpret utterances as concerning either the first or the second element of the centers of centered world, depending on whether it is X or Y who is speaking. The content will then be the same in X's mind as in Y's. This basic idea has been put forward in Ninan 2010 and in Torre 2010.⁴ I shall first discuss Ninan's version, then move on to Torre's.

3.1. Ninan

Ninan first (2010, p. 555) introduces the idea of enriched centers to account for contents that are essentially *de te*, i.e. that essentially concern the addressee *as* addressee, just as a *de se* content essentially concerns the speaker or believer himself. In Ninan's example, if John utters 'Please leave', the content of the utterance is a set of centered worlds $\{\langle x, y, t, w \rangle : y \text{ leaves at } t \text{ in } w\}$.⁵ This is different from the content *that Mary leaves*: $\{\langle x, y, t, w \rangle : \text{Mary leaves at } t \text{ in } w\}$.

As Ninan observes, just enriching the centers with pairs of individuals does not solve the communication problem. For when Mary self-ascribes the content of John's utterance, i.e. $\{\langle x, y, t, w \rangle : y \text{ leaves at } t \text{ in } w\}$, the result is that the belief that *John* is leaving, for from Mary's perspective, the first element of the centers concerns *her*, and the second her interlocutor, John.

The solution Ninan offers is to stabilize the content by making it relative to a *conversational sequence* (Ninan 2010, p. 562). In the basic case we have a pair. Relative to the sequence $\langle \text{John}, \text{Mary} \rangle$, the first element in a center of a centered world applies to John, and the second to Mary. Relative to the sequence $\langle \text{Mary}, \text{John} \rangle$, it is the other way around. Assume the pair $\langle \text{John}, \text{Mary} \rangle$, and the sentence

(3) I like you

as asserted by John. When John relative to the pair $\langle \text{John}, \text{Mary} \rangle$, at time t' in world w' , believes the content

(4) $\{\langle x, y, t, w \rangle : \text{Likes}(x, y, t, w)\}$

he believes the first element, which is John, to like the second, which is Mary, at t' in w' . When Mary believes (4), she also believes the first element to like the second, at t' in w' .

⁴Dirk Kindermann also defended this idea in his presentation at the Barcelona conference in 2012. According to Ninan 2010, n 13, the idea is also suggested in an unpublished manuscript by Seth Yalcin, 2007.

⁵Ninan suppresses the time element.

So, when the contents are the same, and the sequences the same, the resulting beliefs are the same as well.

What is it to have a belief relative to a sequence? One natural interpretation is that relativity to a sequence means *evaluating* the content at that sequence. On this interpretation, believing the content (4) relative to the sequence $\langle \text{John}, \text{Mary} \rangle$ simply is treating $\langle \text{John}, \text{Mary} \rangle$ as part of a point of evaluation. The content is a function from centered worlds to truth values, thereby taking four arguments. Applying that content to the pair $\langle \text{John}, \text{Mary} \rangle$ fixes the values of the first and second variables. The result is a function from the remaining two parameters to truth values, i.e. a function from time-world pairs to truth values:

$$(5) \quad \{ \langle t, w \rangle : \text{Likes}(\text{John}, \text{Mary}, t, w) \}$$

(5) is a *temporal* proposition, not a centered proposition in Lewis's sense, and if we disregard the time parameter, as Ninan does, the result is a classical possible-worlds proposition. Thus, on this understanding of what it is to believe a pair-centered proposition relative to a sequence, it reduces to believing an uncentered proposition. Also, it is not so clear how much is left of the *de se* nature of the belief. Clearly, nothing is left in the content itself. Whatever remains has to do with how the arguments to the function, John and Mary, are selected by the participants, John and Mary. For instance, the sequence might be selected as $\langle \text{me}, \text{my addressee} \rangle$, in which case it is selected in a *de se* way, but as with ordinary Kaplan contents, the way the content was *selected* is not reflected in the content itself. This is contrary to the intentions of Lewis.

Moreover, treating relativity to a conversational sequence in this way bears a strong resemblance to John Perry's conception, for it can be characterized as believing a proposition true *for a conversational sequence*. Formally, the result of fixing the values of the first two of the four parameters is equivalent to *enriching* the original content by adding representations of two possible values, in this case of John and Mary. This is clearly distinct from just applying an unenriched content to the actual point of evaluation in a single step of self-attribution. But how else can the relativity of a content to a conversational sequence be understood?⁶

Ninan's understanding is in fact different. He avoids the result of partial application by understanding in a different way what it is to believe a pair-centered proposition relative to a sequence. According to Ninan (p. 562), that John believes (4) at t' in w' relative to $\langle \text{John}, \text{Mary} \rangle$ amounts to the following:⁷

⁶Thanks here to Manuel García-Carpintero for drawing my attention to the similarity of partial application to Perry's conception of relativity.

⁷This proposal is along the lines Lewis's analysis (1979, p. 539) of what it is to ascribe a property to an

- (6) i) Mary is John's addressee at t' in w' and
 ii) John believes the simple centered-worlds proposition
 $\{\langle x, t, w \rangle : \text{there is a } y \text{ such that } y \text{ is } x\text{'s addressee at } t \text{ in } w \text{ and } x \text{ likes } y \text{ in } w\}$.

Analogously, for Mary to believe (4) at t' in w' , relative to $\langle \text{John}, \text{Mary} \rangle$, is for it to be true that

- (7) i) John is Mary's addressee at t' in w' and
 ii) Mary believes the simple centered-worlds proposition
 $\{\langle y, t, w \rangle : \text{there is a } x \text{ such that } x \text{ is } y\text{'s addressee at } t \text{ in } w \text{ and } x \text{ likes } y \text{ in } w\}$.

Note that in John's centered-worlds proposition, it is the center who likes the addressee, while in Mary's it is the addressee who likes the center. This is the desired outcome.

On this picture, the relation to the pair-centered proposition (4) is not exactly believing.⁸ Rather, we can characterize it as the relation R such that

- (8) $R(a, p, s)$ iff
 i) $a = s_1$ and a believes $E^1(p)$, or
 ii) $a = s_2$ and a believes $E^2(p)$

Here a is an individual, p a pair-centered proposition, and s a pair of individuals. s_i is the i :th member of the pair s . And where p is $\{\langle x, y, t, w \rangle : P(x, y, t, w)\}$, the derived proposition $E^1(p)$ is the simple centered proposition $\{\langle x, t, w \rangle : \exists y(P(x, y, t, w))\}$ that derives from p by existential generalization on the second pair element. Analogously, $E^2(p)$ is the simple centered proposition $\{\langle y, t, w \rangle : \exists x(P(x, y, t, w))\}$, derived from p by generalization on the first pair element.

So, where s is $\langle \text{John}, \text{Mary} \rangle$, and p is the content (4), we have both $R(\text{John}, p, s)$ and $R(\text{Mary}, p, s)$. Consider, by contrast, the content

- (9) $\{\langle x, y, t, w \rangle : \text{Likes}(y, x, t, w)\}$

where the order of the elements x and y are reversed compared with (4). Let q be (9). Let s' be the reversed pair $\langle \text{Mary}, \text{John} \rangle$. Given the beliefs reported in (6ii) and in (7ii), we get the corresponding result that $R(\text{John}, q, s')$ and $R(\text{Mary}, q, s')$. By contrast, it does not hold that $R(\text{John}, p, s')$, or $R(\text{Mary}, p, s')$.

So far, the outcome may look somewhat strange, as if it is relative to a conversational individual. Cf. below, p. 33.

⁸It may be noted here that on this account, if John incorrectly believes that his addressee is Elizabeth, while it is in fact Mary, and (6ii) is true because John believes he likes Elizabeth even though he believes he dislikes Mary, he still comes out as believing (4) relative to $\langle \text{John}, \text{Mary} \rangle$, since his actual addressee is Mary.

sequence what John and Mary believe. This would be strange, since the beliefs that John and Mary have on Ninan's picture are beliefs in simple centered-worlds propositions, and it should be an empirical fact which such propositions they believe, not relative to a choice of a conversational sequence. There is in fact no such relativity, because the *interpretation function* is *also* relative to the sequence,⁹ and this double relativity in fact cancels out.

How this works is explained at Ninan 2010, pp. 563-4. When John asserts

(3) I like you

relative to $\langle \text{John}, \text{Mary} \rangle$, the interpretation is the Kaplanian diagonal of (3). The Kaplan diagonal of a sentence ϕ is

$$\{c : \llbracket \phi \rrbracket^{c,c} = 1\}$$

Contexts here are taken to be simply pair-centered worlds. This means that when John asserts (3), the content is (4). By contrast, according to Ninan, when Mary asserts (3) relative $\langle \text{John}, \text{Mary} \rangle$, the content is the *inverse* of the diagonal. For any pair-centered propositions p, q , q is the inverse of p iff for all pair-centered worlds $\langle x, y, t, w \rangle$,

$$\langle x, y, t, w \rangle \in q \text{ iff } \langle y, x, t, w \rangle \in p$$

(Ninan 2010, p. 564). This means that when Mary asserts (3) relative to $\langle \text{John}, \text{Mary} \rangle$, the content is not (4) but (9). This gives the desired outcomes, in the sense that when the speaker is John, the content is (4), and we have $R(\text{John}, (4), s)$ and $R(\text{Mary}, \text{likes}, s)$, and when Mary is the speaker, the content is (9), and we have $R(\text{John}, (9), s)$ and $R(\text{Mary}, (9), s)$.

This scheme of interpretation in effect means that we have an interpretation function that takes both a sequence and a speaker in that sequence as arguments.

We have the equations

$$\begin{aligned} (10) \quad & \text{i) } \llbracket (3) \rrbracket_s^j = (4) \\ & \text{ii) } \llbracket (3) \rrbracket_s^m = (9) \\ & \text{iii) } \llbracket (3) \rrbracket_{s'}^j = (9) \\ & \text{iv) } \llbracket (3) \rrbracket_{s'}^m = (4) \end{aligned}$$

The subscript provide the sequence argument, and the superscripts, ' j ' and ' m ', provide the speaker, an element in the sequence, John or Mary. Thus, when (3) has the s interpretation, i.e. with respect to the sequence $\langle \text{John}, \text{Mary} \rangle$, and is uttered by John, as

⁹This ingredient in Ninan's theory was not, unfortunately, taken into account in my presentation at the Barcelona conference. Only on rereading the paper did I become clear about its relevance to my discussion.

in (i), the content is (4), and if it has the s' interpretation, with respect to the sequence $\langle \text{Mary}, \text{John} \rangle$, and is uttered by Mary, as in (iv), the content is again (4).

The upshot is that when John's and Mary's simple centered-world beliefs are suitably related, for instance as specified in (6ii) and (7ii), and John (or Mary) assert (3), there are two interpretations that make John and Mary stand in the R relation to the content in question, (4) or (9), and the corresponding sequence.

Does this solve the Lewis communication problem? If the problem was to find interpretations relations between speakers and hearers that make their beliefs relate the right way, the problem is solved. But this is not really solving the communication problem, because it does not by itself show what the speaker John *means*, nor how the hearer Mary *interprets* John's utterance in order to arrive at her belief. In order to add this element, we must assume that the communicators *themselves* apply the interpretation function to succeed in communication. So let's look at the situation again.

If John asserts (3) according to the $\llbracket \cdot \rrbracket_s$ interpretation, what matters is that he is related to what he asserts — (4) — by the $R(\cdot, \cdot, s)$ relation, and if he assert (3) according to the $\llbracket \cdot \rrbracket_{s'}$ interpretation, what matters is that he is related to what he asserts — (9) — by the $R(\cdot, \cdot, s')$ relation. So John must *keep track of* the interpretation he has chosen, and fix on the right conversational sequence, s or s' , that matches the interpretation. Selecting the wrong sequence gives the wrong belief as result. By the same token, if John asserts (3) according to the $\llbracket \cdot \rrbracket_s$ interpretation, what matters to Mary is that she is related to what he asserts — (4) — by the $R(\cdot, \cdot, s)$ relation, and if John assert (3) according to the $\llbracket \cdot \rrbracket_{s'}$ interpretation, what matters to Mary is that she is related to what he asserts — (9) — by the $R(\cdot, \cdot, s')$ relation. So Mary must again *keep track of* the interpretation she has chosen, and fix on the right conversational sequence as well. Selecting the wrong sequence gives the wrong belief as result.

But this means that (10) isn't quite right. Rather, what must come out of the semantic interpretation must be the pair of a pair-centered proposition *and* the corresponding conversational sequence. So, instead of (10), we must have

$$\begin{aligned}
 (11) \quad & \text{i) } \llbracket (3) \rrbracket_s^j = \langle (4), s \rangle \\
 & \text{ii) } \llbracket (3) \rrbracket_s^m = \langle (9), s \rangle \\
 & \text{iii) } \llbracket (3) \rrbracket_{s'}^j = \langle (9), s' \rangle \\
 & \text{iv) } \llbracket (3) \rrbracket_{s'}^m = \langle (4), s' \rangle
 \end{aligned}$$

So, what (iii) indicates is that if Mary chooses the s' interpretation of John's utterance, the result is the pair of (9) and s' . But now we face the question what can mean to interpret an utterance by means of a pair-centered proposition and a sequence, which, except for the

utterance interpretation part is the same as we faced before, of believing a pair-centered proposition relative to a sequence. This time, Ninan offers no answer, and the answer he gave last time is now irrelevant, for both the speaker and the interpreter must somehow have the relevant sequence in mind.

Again, one alternative is to take the proposition to be evaluated with respect to the sequence. Let's follow this alternative. So, suppose that for Mary to interpret John's utterance by means of the s' interpretation amounts to evaluating the pair-centered proposition (9) at the sequence s' . In that case, we get the following result. The evaluation of

$$(9) \quad \{\langle x, y, t, w \rangle : \text{Likes}(y, x, t, w)\}$$

at s' , which is to say at $\langle \text{Mary}, \text{John} \rangle$, is exactly the temporal proposition

$$(12) \quad \{\langle t, w \rangle : \text{John likes Mary at } t \text{ in } w\}$$

Applying this method across the board, the equations in (11) reduce to

$$(13) \quad \begin{array}{ll} \text{i)} \quad \llbracket (3) \rrbracket_s^j & = \quad \{\langle t, w \rangle : \text{John likes Mary at } t \text{ in } w\} \\ \text{ii)} \quad \llbracket (3) \rrbracket_s^m & = \quad \{\langle t, w \rangle : \text{Mary likes John at } t \text{ in } w\} \\ \text{iii)} \quad \llbracket (3) \rrbracket_{s'}^j & = \quad \{\langle t, w \rangle : \text{John likes Mary at } t \text{ in } w\} \\ \text{iv)} \quad \llbracket (3) \rrbracket_{s'}^m & = \quad \{\langle t, w \rangle : \text{Mary likes John at } t \text{ in } w\} \end{array}$$

These are intuitively appropriate results. But we have a peculiar redundancy, since there seem to be two different ways of arriving at the same temporal proposition, depending on the initial choice of sequence. We can without loss remove the redundancy by eliminating the relativity to sequences: by letting 'I' always take the *speaker* of a context of utterance as value, and 'you' always the *addressee*, we don't really need the reversed sequences and the inverses of the diagonals any more, and what remains is simply

$$(14) \quad \begin{array}{ll} \text{i)} \quad \llbracket (3) \rrbracket_s^j & = \quad \{\langle t, w \rangle : \text{John likes Mary at } t \text{ in } w\} \\ \text{ii)} \quad \llbracket (3) \rrbracket_{s'}^m & = \quad \{\langle t, w \rangle : \text{Mary likes John at } t \text{ in } w\} \end{array}$$

On this understanding, the simple centered-worlds beliefs (6ii) and (7ii) that Ninan proposes are superfluous when the interpretation step is taken into account. The evaluation of a pair-centered proposition at a conversational sequence results in a temporal proposition, and if we remove the redundancy, what results is just a notational variant of a temporal proposition theory. The conversational sequence does no real work, since it is determined by the speaker and the addressee of the context.

The alternative is to take the pair interpretations in (11) seriously. The hearer's interpretation really results in a pair of possible-worlds proposition, diagonal or inversely diagonal, and a conversational sequence. Downstream from there, a temporal proposition is generated, and there exist two significantly different ways of getting to the same temporal proposition. I see two problems with this alternative. The first is that of explaining the significance of the pair interpretation. There is supposed to be some important difference between Mary interpreting John as expressing a Kaplanian diagonal relative to one sequence, or the inverse of the diagonal relative to the reversed sequence, but it is not so clear what that important difference is.

The second problem is that, on the assumption that there is such an important difference, what results is a source of possible misunderstanding that will not normally be detected, and perhaps cannot be detected. If John has selected $\langle \text{John}, \text{Mary} \rangle$ in his speaker intention and Mary has selected $\langle \text{Mary}, \text{John} \rangle$, there will be some important miscommunication between them, even if they end up with the same temporal proposition. Lacking an explanation of the significance of the pair interpretation, this is simply implausible. I think the alternative of taking the pair interpretation seriously can be dismissed.

It therefore turns out in the end, with the need of *pairing off* pair-centered propositions with the conversational sequences, that the result is simply a notational variant of temporal, and perhaps in the end, classical, propositions. Solving the Lewis communication problem by means of Ninan's method results in classical propositions in centered-worlds clothing.

3.2. Torre

Stephan Torre (2010) calls his theory of centered worlds with sequences of individuals a theory of "multi-centered worlds" (2010, p. 107). According to Torre, in some situations we need to appeal to (singly or multiply) centered worlds, since it is possible to have beliefs about where one is in the world that don't correspond to any qualitative difference. To illustrate this, Torre uses an example from Lewis, presented as follows:

You have been told that there are two duplicate dungeons. The door of the one on the left leads to freedom. The door of the one on the right leads to the gallows. Open the door and there's no going back. In the duplicate dungeons are two duplicate prisoners. One of them is you. You know, well enough, what sort of world you live in. Narrowing down the alternative worlds won't help you decide whether to open the door. Nevertheless you need to narrow down your alternative possibilities (Lewis 1983a, p. 399).

According to Torre, following Lewis, there is no difference between ordinary possible worlds where (the referent of) "you" are in the dungeon to the left and worlds where (the

referent of) “you” are in the dungeon to the right, unless they differ in other ways. For without any other difference, since the dungeons and the prisoners are each other’s duplicates, there is no qualitative difference between them, and hence no numerical difference either. At least this is so on the *anti-haecceitist* view, and Torre rejects haecceitism (2010, p. 103). For the sake of argument, let us follow Torre here.

Despite this fact about the possible worlds, the prisoner in the dungeon to the left can *believe* that he is the prisoner in the dungeon to the left. The content cannot be a set of standard possible worlds, but must be a set of centered possible worlds. Assume that we name the prisoner that is *actually* in the dungeon to the left ‘L’ and the prisoner who is *actually* in the dungeon to the right ‘R’ (Torre’s naming on p. 106 and again on pp. 111-12). With ‘@’ referring to the actual world, and with time suppressed, the set of centered worlds

$$\{\langle L, @ \rangle, \langle R, @ \rangle\}$$

is the set of relevant centered possibilities for both prisoners: each can be either L or R in the actual world, for all they know (Torre 2010, p. 106).¹⁰

Now Torre adds the following twist to the story of the prisoners:

Let us suppose that you figure out a way of communicating with your fellow prisoner. You learn that he has overheard the guards talking, and he says to you, “You are in the dungeon on the left.” Trusting his information, you open the door and walk towards freedom (Torre 2010, p. 102).

The task is to model the content of this assertion. A simple centered-worlds proposition would give the wrong result: the addressee would arrive at the belief that the *speaker* is in the dungeon to the left. Torre’s solution is to model the content by means of a set of multiply centered worlds. He introduces the terms ‘P1’ for the referent of “you” and ‘P2’ for the referent of “your fellow prisoner” (2010, p. 111), and even adds that the prisoners can have agreed on this naming convention *themselves* (2010, 111, n. 19).

Now Torre assumes available the ordered pair $\langle P1, P2 \rangle$. That is, he assumes that, some way or other, the pair $\langle P1, P2 \rangle$ is chosen for interpretation of indexical sentences, rather than the reversed pair $\langle P2, P1 \rangle$. With this assumption, the content of the utterance by “your” fellow prisoner

(15) You are in the dungeon to the left.

can be given as

(16) $\{\langle x, y, w \rangle : x = L \text{ and } y = R\}$

¹⁰For the sake of uniformity in this paper, I have reordered the elements.

From the assertion of (15), so modeled, and with the ordered pair $\langle P1, P2 \rangle$ given, the addressee, “you”, can infer that P1, i.e. “you”, are L, i.e. the prisoner in the dungeon to the left. “You” then open the door and walk out to freedom. Success. In fact, Torre is here not explicit about how the pronouns are mapped on positions in the sequence. To make the example work, as presented, we must assume that ‘you’ is associated with the first element, and ‘I’ with the second. So we stick to that.

Unlike Ninan, Torre does not make the centered worlds content relative to which element of the sequence $s = \langle P1, P2 \rangle$ is the speaker. So, even with s given, the centered-worlds content of (15) is the same, regardless of whether it is uttered by P1 or by P2. This makes it a crucial condition of communicative success that P1 and P2 have agreed on the same sequence. For suppose that P2 had instead intended the sequence $s' = \langle P2, P1 \rangle$, while P1 had interpreted the utterance with respect to s . In that case, given that ‘I’ selects the second element and ‘you’ the first, P2 would have asserted that P2 is L while P1 is R. P1, on the other hand, interpreting relative to s , would have understood P2 as asserting, as in the original example, that P1 is L, while P2 is R. So, assuming that P2 speaks the truth, in this case P1 walks out and ends up in the gallows. Disaster.

It is thus crucial to Torre’s account that speaker and addressee agree on the proper sequence, i.e. on the order between interlocutors. The question is how they can do that. It would not be of great help for P2 to say

(17) I am the second element, you are the first.

For since (17) itself contains the pronouns, it must already be interpreted relative a sequence, and so the problem must already have been solved.

The alternative method is for the speaker to manage to refer to himself or his fellow prisoner by some other method. For, instance, he may try

(18) The speaker of this utterance is the first element, the addressee the second.¹¹

We can immediately observe that if this method of identifying the speaker is available, it can be used directly:

(19) The speaker of this utterance is in the dungeon to the right.

Given that we have left time out of consideration, the content of (19), as well as the content of (17), is a non-centered proposition. But non-centered propositions were supposed not to be available to distinguish between the two possibilities, that P1 is L and

¹¹In fact, pretty much this formulation was proposed by Manuel García-Carpintero on Torre’s behalf during the Barcelona conference, and met with approval by Torre.

that P1 is R, because of the symmetry. We were told by Torre that whether P1 is L and P2 is R, or whether P2 is L and P1 is R, we still have the same ordinary possible world, and so no non-centered possible-worlds proposition could distinguish between these two possibilities. Therefore, (19) was supposed not to be available as conveying a communicable content that distinguishes between them. But then, (18) is not available either, since it relies on exactly the same method of identifying the speaker.

In fact, it is easy to see that the problem is completely general. P1 and P2 cannot use *de se* communication to agree on which sequence to use, *s* or *s'*, since that presupposes the communication problem to be solved, and they cannot use *de dicto* communication for this purpose either, since that presupposes that there is at least *some* asymmetry between the prisoners that can be exploited in designating one of them rather than the other, and by assumption they are duplicates in duplicate dungeons, so there isn't any. Hence, the communication problem actually cannot be solved by Torre's method. The assumption that the prisoners can use something like (17) to solve it actually is an assumption that the communication itself creates an asymmetry, after which the prisoners are no longer duplicates. If they are not, by Torre's own lights (2010, p. 101), and as exemplified by (19), centered contents are not needed.

4. Associated uncentered contents

The main idea of appealing to associated uncentered contents, or *de dicto* contents, is that what the speaker communicates is a *de dicto* content that is suitably *related* to the *de se* content the speaker believes. The believed content is modeled as centered content, while the communicated content is modeled as uncentered. Still, on learning the communicated content, the hearer can form an associated *de se* content himself, that suitably corresponds to the speaker's *de se* content. In this way, centered contents can be indirectly communicated, by way of a detour over *de dicto* contents. This method has been proposed by Sarah Moss (2012) and by Max Kölbel (2013). I shall first look at Moss's version, then Kölbel's.

4.1. Moss

According to Moss, for any centered content, given what the subject *believes with certainty*, there is always a *de dicto* content that is true in just the same belief-worlds. This claim is stated in the following principle (Moss 2012, p. 226):

(PROXY) Given a *de se* proposition, there is a *de dicto* proposition such that for any centered world compatible with what you believe, that centered world is in the former proposition just in case it is in the latter.

In Moss's terminology, a *de dicto* proposition is a set of centered worlds where only the world component matters. That is, if P is a *de dicto* proposition, then if $\langle x, t, w \rangle$ is P , so is $\langle y, t', w \rangle$, for any individual y and time t' .¹²

Prima facie, (PROXY) is a strong thesis and looks implausible, but according to Moss, there are always suitable ways of thinking about oneself, or about a time, such that the thesis comes out true. She illustrates this with a story about David Kaplan:

[...] suppose Kaplan has gotten lost in his sea kayak and calls you for directions. He describes the coastline and you look at maps, but you still cannot figure out where he is paddling. There are two possible worlds compatible with your beliefs as you talk to Kaplan: one where Kaplan is in Bellingham, and one where Kaplan is in Seattle. Just the same possible worlds are compatible with what Kaplan believes. Neither you nor Kaplan can figure out which of the worlds is actual. And so Kaplan is not only ignorant of *de se* facts. He is also ignorant of the same *de dicto* facts as you. Suppose Kaplan is in Bellingham and someone exactly like him is in Seattle, and suppose Kaplan introduces a name for himself as he is talking to you

(3*) Let 'Dr. Demonstrative' name myself.

In one possible world compatible with what you and Kaplan believe, Dr. Demonstrative is in Bellingham and some other guy is in Seattle. In another world compatible with your beliefs, Dr. Demonstrative is in Seattle and some other guy is in Bellingham. Each possible world corresponds to two centered worlds: one centered on Bellingham and one centered on Seattle. Since Kaplan believes that he himself is Dr. Demonstrative, he can rule out exactly two of these four centered worlds. That is why his *de se* belief that he himself is in Bellingham is equivalent with his *de dicto* belief that Dr. Demonstrative is in Bellingham, given what he believes. The same goes for all of his *de se* attitudes (Moss 2012, p. 228, asterisk added).

To spell this out in detail, let Y be (the referent of) "you", D Dr. Demonstrative, O the other guy (the relevant witness of "some other guy"), B Bellingham, S Seattle, and I the located-in relation. We have two ordinary worlds: w_b , where D (Kaplan) is in Bellingham and O in Seattle, and w_s , where D is in Seattle and O in Bellingham. Suppressing the time element, the *de se* proposition that is the content of the belief to be in Bellingham is

$$(20) \quad \{\langle x, w \rangle : I(x, B)\}$$

The *de dicto* proposition that D is in Bellingham and O in Seattle is

$$(21) \quad \{\langle x, w \rangle : I(D, B) \& I(O, S)\}$$

¹²Moss 2012, p. 226. To make this paper uniform, I have reordered the variables.

This is a *de dicto* proposition in Moss's sense, since it places no condition on the center individual.

Now we add the centered worlds, two for each ordinary world. We have the following: $\langle D, w_b \rangle, \langle O, w_b \rangle, \langle D, w_s \rangle, \langle O, w_s \rangle$. (21), the content of Kaplan's and *Y's* *de dicto* belief that Dr. Demonstrative is in Bellingham, contains the first two: $\langle D, w_b \rangle$ and $\langle O, w_b \rangle$.

As regards the content of Kaplan's *de se* belief that *he* is in Bellingham, which is (20), any centered world included in its content must be such that the center is in Bellingham. This rules out $\langle D, w_s \rangle$, where Demonstrative is the center but in Seattle, and $\langle O, w_b \rangle$, where Other Guy is the center but in Seattle. The two remaining worlds, $\langle D, w_b \rangle$ and $\langle O, w_s \rangle$, are included:

$$(22) \quad \{\langle x, w \rangle : I(x, B)\} = \{\langle D, w_b \rangle, \langle O, w_s \rangle\}$$

According to Moss, we must now take into account what Kaplan believes with certainty, and that includes Kaplan's certain belief that he is Dr. Demonstrative. This is the belief

$$(23) \quad \{\langle x, w \rangle : x = D\}$$

The centered worlds included in this set are all and only worlds with Demonstrative as center:

$$(24) \quad \{\langle x, w \rangle : x = D\} = \{\langle D, w_s \rangle, \langle D, w_b \rangle\}$$

The other two centered worlds are incompatible with what Kaplan believes. Restricting Kaplan's *de se* belief that he is in Bellingham to worlds compatible with what he believes, that he is Demonstrative, then amounts to intersecting the *de se* belief, (20) with the certain belief, (23). The result is $\{\langle D, w_b \rangle\}$.

The *de dicto* belief (21) that Demonstrative is in Bellingham is

$$(25) \quad \{\langle x, w \rangle : I(D, B) \& I(O, S)\} = \{\langle D, w_b \rangle, \langle O, w_b \rangle\}$$

We must now also restrict the *de dicto* belief to world compatible with what Kaplan believes with certainty, which means intersecting (25) with (24). The result is again $\{\langle D, w_b \rangle\}$. This is in accordance with (PROXY). Every world compatible with what Kaplan believes with certainty, i.e. (23), is in (20) iff it is in (21), for $\langle D, w_b \rangle$ is in both and $\langle D, w_s \rangle$ is in neither.

On the exemplified interpretation of (PROXY), it amounts to the following:¹³ for any

¹³Discussion of a first draft with Manuel García-Carpintero helped me get clearer about Moss's ideas.

de se belief *G*, there is a *de dicto* belief *H* and a certain belief *J* such that

$$G \cap J = H \cap J.$$

On this interpretation, the believable subset of a non-communicable *de se* belief coincides with the believable subset of communicable *de dicto* belief. So he can then sincerely assert the latter iff he believes the former, given his certain beliefs. Prima facie, this looks like a method that makes progress with the communication problem, but as far as I can see, it cannot succeed.¹⁴

In Moss's example, Kaplan's certain belief, that *he* is Dr. Demonstrative, is itself a *de se* belief. In order to successfully communicate his *de dicto* belief that Dr. Demonstrative is in Bellingham, he needs to inform his interlocutor, *Y*, about who Demonstrative is, and this is done by means of the stipulation (3*):

(3*) Let 'Dr. Demonstrative' name myself.

The problem with this is immediate: since it contains the (unbound) personal pronoun 'myself', an utterance of (3*) itself has a *de se* content. Hence, using this stipulation presupposes that the problem of communicating *de se* propositions has already been solved. As in the case of Torre, any method that can be applied to solve it in this particular case, for the stipulation about the name, can be used directly for conveying *de se* belief, bypassing the need for this special method. So, unless there is another method, the use of the stipulation (3*) will fail. By assumption, there is no other method.¹⁵

Moss uses this example for convincing the reader that there is a method that works, and the example fails. Can the method still work, perhaps with a different implementation? There are two features of the example that create the problem when combined: The first is that the certain belief *J* that does the intersecting, in this case the belief of being identical with Demonstrative, is itself *de se*, and the second is that this belief, in this case

¹⁴A minor problem is that Moss requires a restriction to what the subject believes with certainty, not that this belief be true. If Kaplan has a certain belief that he is the Other Guy, he can assert a *de dicto* proposition that satisfies (PROXY), that *the Other Guy is in Bellingham*, but this proposition is false although his *de se* belief is true. Of course, in this case, the intersection of his *de se* belief with his deluded identity belief is empty, and so is the intersection of his *de dicto* belief with this deluded certain belief. But (PROXY) does not require the intersection to be non-empty.

The certain belief that Moss proposes, the *de se* belief to be Dr Demonstrative, is not only believed with certainty but also a priori true, since true by definition. So, this minor problem is easily solved by requiring both certain belief and truth, or perhaps a priori knowledge. In fact, this is probably what Moss does have in mind, since she appears to want to exclude from certainty anything for which there *can* be "some shred of doubt" (Moss 2012, p. 229), and so she most probably does not mean merely subjective certainty.

¹⁵What about "The speaker of this utterance"? It is possible that the evil demon deludes the speaker into believing that he is the speaker of an utterance that is really made by someone else, and hence possible to doubt that one is the speaker of an utterance one in fact makes.

(3*), or a belief equivalent to it, is itself communicated. Are both features necessary?

The first is indeed necessary. The result of intersecting the *de dicto* proposition H with J must be a *de se* proposition, for it must coincide with a *de se* proposition, a subset of G . Because of this, J itself must be *de se*, for intersecting a *de dicto* proposition with a *de dicto* proposition yields another *de dicto* proposition: a centered world $\langle x, w \rangle$ is in the intersection of two *de dicto* propositions only if $\langle y, w \rangle$ is as well, for any available center y , because by definition, if $\langle x, w \rangle$ is included in a *de dicto* proposition, so is $\langle y, w \rangle$. So, as in Moss's example, the certain belief J must be *de se*.

Must it also be communicated? In Moss's example, the need of communicating (3*) depends on the fact that it is a stipulation about the name-using. But the stipulation is motivated by the need of certainty. As Moss says (2012, p. 229), Kaplan can doubt that his pants are on fire and that he is David Kaplan, but he can't doubt a stipulation. Can anything else than a stipulation meet the requirement?

Here is an argument that nothing else can: The content, in order to be *de se*, must be of the form $\Phi(x)$, for some condition Φ . In order that the condition's holding of x be indubitable by the subject, it must be analytic in nature (e.g. $\lambda z(z = z)(x)$). But in order that it hold *uniquely* of x , and be analytic, it must be stipulated to hold of x .¹⁶ Any condition that is not stipulated to hold uniquely of x might, for all x can believe without doubt, hold of someone else.¹⁷

Then, precisely because the condition holds of the subject by stipulation, it must be communicated to be known by anyone else. And this leads us back to square one.

The main topic of Moss's paper is belief updating.¹⁸ She uses communication as a model for updating: the subject communicates earlier beliefs to her later self: "In hypothetical black box updating, you form beliefs on the basis of information you get from your previous self. Getting information from your previous self is just like getting information from other agents." (Moss 2012, p. 238). It is possible that the problems I have stressed above for her theory of communication do not arise, or can be avoided, in the special case of communicating with one's later self. My interest here, on the other hand, is the general case of communication, and the question whether Moss's proposed method does offer a solution to the problem of communicating centered *de se* contents, whether or not that was also her interest.

¹⁶It doesn't have to be a proper name. It can also be a predicate, as in "Let 'Demonstrativizes' be true exactly of me". Then use the description 'the Demonstrativizer'.

¹⁷The editors have stressed that it is also indubitable to Kaplan that *he exists*, and that *he is thinking*. Agreed. But it cannot be indubitable to Kaplan that *he is the only thinker*, even if it is true. Moreover, if he uses the description 'the thinker' in communication with someone else, the *de se* proposition that *he is the only thinker* is anyway not true either.

¹⁸Both the editors and an anonymous referee have suggested that because of this, my remarks on her treatment of the communication problem are unfair. Perhaps they are right.

4.2. Kölbel

Max Kölbel (2013) also pursues the idea that communication uses *ersatz* non-*de se* contents to convey information, when needed. As we shall see, Kölbel's idea is very close to Frege's.

Kölbel uses the abstract idea of a *location*, which plays the role of the center in centered worlds, and does not commit himself to any particular version of centered contents. In briefly presenting his general ideas of about content and information, I shall stick to this notion of a location, and then switch to the standard idea of individual-time pairs, while suppressing the time part.

Kölbel defines the concepts of *portable* (2013, p. 100) and *locally portable* contents, or *portable relative to a class of locations* (2013, p. 102) as follows. Here W is the domain of possible worlds, L the domain of *locations*, and P^C the domain of centered-worlds propositions:

(P) For all $p \in P^C$: p is *portable* iff for all $w \in W$ and for all $l_1, l_2 \in L$ ($\langle l_1, w \rangle \in p$ iff $\langle l_2, w \rangle \in p$).

(RP) For all $p \in P^C$: p is *portable relative to a class* $C \subseteq L$ iff for all $w \in W$ and for all $l_1, l_2 \in C$ ($\langle l_1, w \rangle \in p$ iff $\langle l_2, w \rangle \in p$).

Formally, these conditions concern locations with respect to every possible world. There is question whether locations *exist* in all possible worlds. Concrete individuals do not exist in all worlds. If space-time locations are *relative*, then space-time locations do not exist in all worlds either. In particular, a spatial location on Earth does not exist in possible worlds where Earth itself does not exist.

It is then natural to require that a pair $\langle l, w \rangle$ exists only if l exists in w . In that case, the quantification over worlds should also be restricted in (P) and (RP) to worlds where both l_1 and l_2 exist. The alternative is to say that the pair $\langle l, w \rangle$ does exist even if l does not exist in w , but that this pair is not included in any proposition p whose truth at a centered world depends on the center. Let's call the first alternative the *world-restricted interpretation*, and the second the *world-unrestricted interpretation*.¹⁹

The interest of local portability is that some *de se* contents need not be portable, but may be true of every participant in a particular conversation (an idea similar to ideas in Egan 2007). With the quantification over worlds, however, the requirement is strengthened to that of being *necessarily* true of all or none in the conversation. This is especially

¹⁹I had originally tacitly assumed the world-restricted interpretation, and a proposed counterexample below depends on it. An objection by the editors against the proposed counterexample made me realize the need for this distinction.

strict if we don't restrict worlds to those where all participants in the conversation exist.

Corresponding to these concepts, Kölbel defines his notion of ersatz contents, which he calls *portable surrogates*. This is defined by way of *correctness of belief* (2013, p. 103):

- (C) For all $w \in W$, for all $p \in P^C$ and for all $l \in L$, it is *correct to believe* p at l in w iff $\langle l, w \rangle \in p$.

Then the idea of a *portable surrogate* is introduced (2013, p. 104):

- (PS) For all $p, q \in P^C$ and $l \in L$, q is a *portable surrogate of* p at l iff: for all $w \in W$, it is correct at w to believe p at l iff for all $l^* \in L$ it is correct at w to believe q at l^* .

Finally, there is the corresponding idea (2013, p. 107) of a *locally portable surrogate*:

- (LPS) For all $p, q \in P^C$, for all $l \in L$ and all $C \subseteq L$, q is a *portable surrogate of* p at l relative to C iff: for all $w \in W$, it is correct at w to believe p at l iff for all $l^* \in C$ it is correct at w to believe q at l^* .

Switching over to just using individual centers as location, the idea of portable surrogate can be easily illustrated. Consider

- (26) a. I am hungry.
b. K is hungry.

and assume (26a) believed by K . The contents will be equivalent in the following sense:

- (27) It is correct of K in w to believe *de se* that *he* is hungry iff it is correct of anyone in w to believe that K is hungry.

Kölbel provides a short derivation (2013, p. 7) to show that this equivalence holds abstractly and in general. Schematically, we have, corresponding to (26), the contents,

- (28) a. $\{\langle x, w \rangle : H(x, w)\}$
b. $\{\langle x, w \rangle : H(K, w)\}$

with H for 'is hungry'. It is easy to see that the *intersection* of (28a) and (28b) includes

- (28) c. $\{\langle x, w \rangle : x = K \& H(x, w)\}$

It also includes $\langle J, w \rangle$, provided both J and K are hungry in w . Nevertheless, we can see that the following is true:

- (29) For all w : for all x ($\langle x, w \rangle \in (28b)$) iff $\langle K, w \rangle \in (28c)$ iff $\langle K, w \rangle \in (28a)$.

Thus, given how *correctness of belief* is defined, bringing it in is a bit of a detour.

Kölbel differs from Moss in not attempting to generate some *de dicto*, or portable surrogate, that *coincides* with the *de se* content in some restricted domain. Rather, what matters for the portable surrogate is only that the set of *worlds* is right, and the relevant worlds can be extracted from the *de se* contents (28b) and (28c).

According to Kölbel, only portable contents matter for information transfer:

Successful information transfer occurs when, as a result of witnessing the linguistic act, a recipient comes to believe a surrogate of the content believed by the source at the source's location (Kölbel 2013, pp. 105-6).

Since only the corresponding classical proposition matters for the portable surrogate, the question is what we need the centered contents for. Kölbel considers this question:

First, proponents of a framework of purely portable contents (e.g. Frege, Perry, and many others) will object that something like the surrogate contents I am mentioning can and should figure as the contents of beliefs (and assertions) in the first place. Then we would not need a theory of portable surrogates, and could simply stick to the familiar model in which information transfer is simply content transfer. Therefore, a theory of non-portable contents that postulates informationally equivalent portable surrogates must be inferior to a theory that avoids non-portable contents in the first place. Not so. Centered contents were introduced to do a certain job, and they still need to do that job, even if, when explaining information transfer, we resort to portable surrogates (Kölbel 2013, p. 106).

This can be taken in two ways. On one way of thinking, we might say that centered *de se* contents are needed for contents of belief, while only portable surrogates (or classical propositions) have any role at all to play in communication, or else that somehow, centered *de se* contents are relevant for communication as well, even though we have to switch to surrogates for explaining information transfer. The second is the stronger one, but strikes me as incorrect.²⁰

If only surrogates are relevant for communication, then only surrogates are relevant to semantics to the extent that semantics is relevant to communication. The question

²⁰I do think the first is incorrect as well, although I'll not argue that at any length in this paper. Cf. section 6.

On the other hand, centered contents are not sufficient. Lewis appealed to centered contents, among other things, in order to account for how the deluded Heimson can believe that he is Hume (Lewis 1979, pp. 536-8). This is needed, for there is no world where Heimson is Hume, whereas the content $\{\langle x, w \rangle : x = \text{Hume}\}$ can be self-ascribed by both Hume and Heimson, truly by the former, falsely by the latter. However, Heimson can be deluded enough to believe that he is both Hume and Reid, given that he believes that Hume is identical to Reid. There is no centered world where Hume is identical to Reid, so there is no centered-propositions account of the possibility of this belief anyway.

is whether *de se* contents, non-portable or locally portable, are ever relevant semantic values.

In the latter part of his paper, Kölbel considers how surrogates get expressed by sentences, in order to explain information transfer by means of language. Kölbel introduces a distinction between the *semantic value*, SV, of a sentence, and its *conversational content*, or CC (2013, p. 14). The presentation suggests that conversational content is something distinct from semantics, but there are good reasons to regard it rather as a distinct semantic *function*, for instance as part of a semantic system that includes both SV and CC as meanings (however, this could be regarded as a terminological issue regarding ‘semantics’).

Kölbel illustrates the distinction by the following pair of contents:

- (30) a. $\{\langle x, w \rangle : x = \text{Lauben}\}$
 b. $\{\langle x, w \rangle : \text{the speaker of utterance } u = \text{Lauben}\}$

Here utterance *u* is the utterance produced by the speaker, which can be made explicit by means using ‘this utterance’ as a complex demonstrative. The content (30b) is then not *de se*, but any world where Lauben actually produces this utterance is included in both (30b), with any center, and in (30a), with Lauben as center.

As Kölbel introduces it, CC agrees with SV for a sentence *s* except if *s* contains a “CC-modifying expression”. If the SV of *s* is a non-portable content, then the CC of *s* will be a *locally portable surrogate*, whose extension will differ from that of the SV value in some contexts. On the Stalnakerian conversational model that Kölbel subscribes to, it is the CC-value of *s* in the context that get’s added to the common ground, and hence that is relevant to communication (2013, p. 15). The question is then, firstly, whether any non-portable values are ever relevant, and, secondly, whether any locally but not globally portable contents are ever relevant.

Kölbel proposes the pure indexicals as CC-modifying expressions. The reader is not provided with the mechanics of the modification, but it seems that Kölbel in general has a token-reflexive interpretation of the indexicals in mind. This means that (31a) will be interpreted as (31b)

- (31) a. I am Lauben.
 b. The speaker of this utterance is Lauben.

(these two sentences will embed differently in linguistic contexts, so some further apparatus is needed to take care of this). This means that the communicated content of (31a) is (30b) rather than (30a).

It is not so clear in what sense this interpretation is a *modification* of anything. It simply looks as if the CC-interpretation of ‘I’ is the token-reflexive interpretation presented in (31b), and that is the interpretation relevant for communication, whether one calls it “semantic” or not. If this were the end of the story, we should say that on Kölbel’s picture, centered content play no role at all for communication, even if they are important for the attitudes.

It is not the end of the story, however, for two reasons: on the one hand, Kölbel thinks there are sentences which do get a non-portable interpretation by CC, and on the other, the CC surrogates need be only locally portable. Regarding the first point, Kölbel proposes (inter alia) the following two sentence pairs (2013, p. 116):

- (32) a. It’s raining.
b. It’s raining *here*.
- (33) a. The “living statue” at the end of Ramblas is moving.
b. The “living statue” at the end of Ramblas is moving *now*.

According to Kölbel, the a-sentences are non-portable or locally portable, depending on the conversational context (whether the conversation participants are at the same place or not ((32a)), or take part at the same time not ((33a))), while the b-sentences are guaranteed to be locally portable.

As regards (33), the difference between (a) and (b) is analogous (apart from idea of *local* as opposed to global portability) to the difference between them in Kaplan’s semantics. This concerns *temporalism* about propositions, which I here set aside as distinct from centered-worlds propositions. The topic is too large to be covered here.

As regards (32a), on the other hand, I find it simply incredible that it could ever be used to assert a non-portable centered-worlds proposition such that: a) it is true as asserted at time *t* by Max in Barcelona, since it is raining in Barcelona at *t*, b) understood and believed at *t* by his addressee Isidora in Paris, and c) falsely so, since it is *not* raining in Paris at *t*. That would be the centered-worlds content, on a model where spatial locations are element in centers (as on Lewis’s model), and as far as I can judge, neither (32a), nor any other sentence is ever used that way.

It can be replied here,²¹ that the use of ‘It is raining’ as expressing a centered content can indeed go wrong, as exemplified, but that it should be restricted to contexts where it does *not* go wrong, which means to contexts where the participants of the conversation are all at the same location. However, as regards the actual use of ‘It is raining’, there simply is no such restriction. It is often used in conversations where participants are not

²¹It was actually replied by Kölbel during a seminar in Barcelona.

at the same location.²²

Turning to locally portable uses of 'I', suppose Max utters (26a) to Isidora

(26a) I am hungry.

and both of them have *de se* beliefs with content (28a)

(28a) $\{\langle x, w \rangle : H(x, w)\}$,

then both have correct beliefs if both are hungry at the time. This is not, however, enough for making the belief locally portable in the Max-Isidora conversation, for the definition (RP) of *portability relative to a class of locations* requires the two locations—in this case the two individuals Max and Isidora—to be centers in exactly the *same* centered worlds within the set (28a). Since their states of hunger are not necessarily correlated, this condition is not met.

Still, it is not hard to construct examples where the condition is met. Suppose the thesis of the necessity of origin is true: in every world where human A exists, the father/mother of A exists as well. Assume, further, the *world-restricted* interpretation. Then, assuming Max and Isidora are siblings, and Max truly utters

(34) Silvia is my mother

Isidora, believing what Max says, would come to believe that Silvia is *her* mother, for on Kölbel's account, 'my' is here interpreted as locally portable, and it suffices for local portability that the property of being Silvia's offspring is shared in every world where Max and Isidora both exist (which is all we need care about on the world-restricted interpretation). But this is clearly not an intuitively available interpretation of (34).

Assume instead the *world-unrestricted* interpretation. In that case, (34) does not go through as a counterexample, since there are worlds where Max exists but Isidora does not. So the content is not then portable relative to the set {Max, Isidora} of locations. But then, consider instead

(35) I am a member of the Max-Isidora duo.

Whether said by Max or by Isidora, it is true only in worlds where both Max and Isidora exist. So, it is locally portable relative to {Max, Isidora}. So, Max's utterance should be believed by Isidora, on Kölbel's account, in the sense that *she* is a member of that duo,

²²It may be added here that this use of wide locations, capable of housing several individuals, is very different from Lewis's use of time-location pairs as centers of centered worlds. Lewisian locations can only house a single individual.

but again, this is counterintuitive. On top, it is strongly counterintuitive that the interpretation of the pronoun *depends* on the property ascribed to its referent.²³

The conclusion is that those ingredients of Kölbel's views that give centered contents *some* role to play in communication (beyond being contents of attitudes that *motivate* communication), simply are not intuitively acceptable.

5. Content transformations: recentering models

The last approach that I'll consider uses the idea that the hearer's interpretation of the speaker's utterance involves a so-called *recentering*. This is a content-transformation. The hearer *transforms* the centered content that fits the speaker-center so that the transformed content (the transform) fits the hearer, and is true of the hearer iff the original content is true of the speaker. That is, when a speaker ascribes a property *F* to herself, the hearer will self-ascribe a corresponding property *G*, the property that *her interlocutor* has property *F*. This line has been taken by Allan Gibbard (2012) and Clas Weber (2013). They both use the term 'recentering'. I shall devote less attention to Gibbard than to Weber, since Gibbard's account is less worked out, and also does not in fact fully employ the centered-worlds framework.

5.1. Gibbard

In the Appendix 1 of his 2012, Gibbard argues against the idea that classical propositions are the objects of belief, as well as the idea that they are transferred in communication. Gibbard uses a two-dimensional framework. We have a vertical dimension of possible worlds and a horizontal dimension. The worlds in the vertical dimension are the worlds of utterance of a sentence, which are also the worlds that provide the perspective of the thinker. The worlds on the horizontal dimension are the worlds of evaluation. Together, for each sentence *s*, we get a matrix, where each cell corresponds to a pair $\langle w, w' \rangle$ of worlds, where *w* is a world of utterance, and *w'* a world of evaluation. The truth value in this cell is the truth value the proposition *s* expresses in *w* as evaluated in *w'*. Following Kaplan (1989), Gibbard calls the matrix the *character* of the sentence.

What he calls *propositions* are rows in such matrices. What Gibbard calls *doxons*, on the other hand (2012, p. 261), are the *diagonals* of the matrices, following the idea in Stalnaker (1978). The diagonal of a sentence *s* is the set of worlds *w* such that *s* is true at $\langle w, w \rangle$. What is transferred-and-transformed in communication, according to Gibbard, is the doxon. And the doxon, to be understood, requires recentering. To this speaker-time pairs are added, to get centered worlds. To exemplify (Gibbard 2012, p. 260), when

²³Kölbel also proposes examples involving predicates of personal taste, and generic 'one'. There is no space here for discussing those examples. As regards the latter, cf. Stanley 2011, Chapt. 3.

Ann writes in a letter

(36) I am happy now

the character of the sentence is such that it is true on the diagonal, at a centered world $\langle x, t, w \rangle$ iff x at t is happy in w . This is seen as a property, the property of the speaker-time-world to be happy-at-the-time in the world of utterance. Ann, at time t , self-ascribes this property by means of (36). When Ben reads the letter, he needs to characterize *from his standpoint* how things are with Ann, e.g. by accepting

(37) She was happy then

In thought, Ben then performs a recentering. However, described so far, everyone agrees with what has been said. In order to preserve content, the hearer needs to change the pronouns. What the speaker expresses with 'I', in face-to-face conversation, the hearer has to express with 'you', or something contextually equivalent. The question is what recentering amounts to when we don't have access to representations of the centers, i.e. of the subject and the time. In the centered-worlds proposition model, we don't have access to these representations.

Unfortunately, Gibbard seems not to have taken this requirement completely to heart. For according to Gibbard, the recentering takes place at the level of *structured* thoughts, where the representations of the center, individual and time, is included:

What writer Ann voices with 'I' reader Ben thus transforms into a concept he voices with 'that woman, and he recenters this transformed concept on himself. Likewise with time: Ben's word 'then' voices, in this utterance, the concept of whenever this letter was written, again taken rigidly. What Ann voiced with 'now' and present tense Ben transforms into what he voices with 'then', and he centers this on himself now (Gibbard 2012, p. 267).

Element by element transformation with recentering, after all, preserves reference, and the structured proposition is a structure composed of a relation and these referents—the referents being Ann and the time of her writing the letter (Gibbard 2012, p. 267).

Gibbard's model thus makes representations of oneself and the current time available. The structured thought is what is primarily expressed and transferred to the hearer. The structured proposition is transformed by the hearer into another structured proposition, where the speaker time and individual are represented from the hearer's perspective. In the final step, the hearer goes on to a centered-worlds proposition, derived from the

transform, and centered on himself at the time of hearing or reading. At least, this is the interpretation of Gibbard that I managed to extract from the text.

The model makes use of what was supposed not to be available in the centered contents theory. The problems of *de se* beliefs that were supposed to be solved by appeal to centered-worlds propositions therefore recur in pretty much their original form. Weber's version of the recentering idea treats the theory of centered contents more consistently.

5.2. Weber

Weber characterizes what he calls "The FedEx Model" of communication by the following two principles (2013, p. 208):

The mind-to-speech principle: the content of the utterance is identical to the content of a belief the speaker expresses.

The speech-to-mind Principle: the content of the utterance is identical to the content of a belief the hearer acquires.

Because of the communication problem, both principles cannot be accepted together with centered contents (and the assumption that beliefs can be expressed; cf. 2013, p. 209), and Weber prefers to reject the second.

A crucial ingredient in the model is the hearer's belief about his relation to the speaker. Whatever else the hearer may believe, if there is communication at all with a particular pair of speaker and hearer, the hearer can always locate the speaker by means of the relation *R* (2013, p. 209):

(R) *h* and *s* stand in *R* just in case *s* produced the utterance token that *h* is perceiving.

With this relation defined, Weber's model is presented as follows (2013, p. 212):

The recentering model

1. The hearer perceives an utterance "u" [Perceiving]
2. The hearer believes that the expressed content of "u" is true of the speaker. [Centering]
3. The hearer believes that she is *R*-related to the speaker. [Locating]
4. The hearer infers information about herself from 2. and 3. [Recentering]

To complete the picture, Weber characterizes *centering* and *recentering* as follows (2013, p. 213):

Centering

1. *Understanding premise:* "u" is true of the set of Φ -individuals.

2. *Trusting premise*: “u” is true of the speaker.
3. *Centering conclusion*: The speaker is a Φ -individual.

Recentering

1. *Centering premise*: The speaker is a Φ -individual.
2. *Locating premise*: I am *R*-related to the speaker.
3. *Recentering conclusion*: I am *R*-related to a Φ -individual.

The locution “is true of the set of Φ -individuals” must here be understood in the sense that the utterance is true of all and only Φ -individuals, rather than as true of the set. Weber provides two examples, the first of which runs as follows:²⁴

(38) **Weber’s Example 1** (2013, pp. 213-4)

1. B perceives the utterance “You have Groat’s disease” (being produced at the speaker’s spatio-temporal location) [Perceiving]
2. B understands “You have Groat’s disease”: “u” is true of exactly those individuals whose addressee has Groat’s disease. [Understanding]
3. B believes that “You have Groat’s disease” is true of the speaker. [Trusting]
4. B believes that the addressee of the speaker has Groat’s disease. [Centering]
5. B believes that she is the addressee of the speaker. [Locating]
6. B infers information about herself from 4. and 5.: she has Groat’s disease. [Recentering]

Here steps (2)–(4) comprise the Centering stage, while steps (4)–(6) comprise the Recentering stage.

At a first glance, the example makes perfect sense: this is exactly how a hearer might arrive at a belief about herself from an utterance by a speaker. However, this presupposes that she believes the speaker. So the Trusting step, (3), is essential. This cannot be in general required, however, or we could not make sense of exchanges like the following:

- (39) A: You have Groat’s disease.
 B: I don’t believe that. / That’s not true.

In communication we typically first *understand* what the content of a speaker’s utterance is, and then assess it as true or false with respect to the actual evaluation point (cf. Section 2). This is not just subscribing the so-called “FedEx” model which Weber rejects, for it does not presuppose that the content the hearer assess is the *same* as what the speaker

²⁴I have abridged it by not repeating Centering, and by changing formulations, omitting superfluous mention of sets.

thinks or expresses. It only presupposes that the hearer grasps *the relevant content* in advance of taking a stand to its truth value. This is also what is required to make sense of a dialogue like (39).

However, on a centered contents model, there is no centered content that both concerns the subject from a *de se* perspective *and* is only entertained, not believed. This is so, since on Lewis's model, it is only insofar as a centered content is *self-ascribed* by a subject that a property is predicated of the subject at all. The predication of the property is *not* part of the *content*; it enters only in the self-ascription step, which is the centered-content counterpart to ordinary belief. Without going *non-centered*, and adding a representation of the subject in the content *itself*, the subject cannot grasp without belief what is predicated of her. This shortcoming does not matter as long as we are only concerned with belief itself, as Lewis was, but in communication we come across contents that we may believe or not.²⁵

To see how this plays out on Weber's account, we have to see how far the hearer gets without relying on the Trusting step. Since the Centering step depends on the Trusting step, without the latter, the former is not available either. Below I shall add an Extended Understanding step, to get rid of the meta-linguistic representation:

(40) **Weber's Example 1 without Trusting**

1. B perceives the utterance by A of "You have Groat's disease" (being produced at the speaker's spatio-temporal location) [Perceiving]
2. B understands "You have Groat's disease": "u" is true of exactly those individuals whose addressee has Groat's disease. [Understanding]
3. B understands "You have Groat's disease": B understands that A self-ascribes the property of having an addressee with Groat's disease. [Extended Understanding]
4. B believes that she is the addressee of A, i.e. self-ascribes the property of being the addressee of A. [Locating]
5. B self-ascribes the property of being the addressee of a speaker who self-ascribes the property of having an addressee with Groat's disease. (From (3) and (4)).

In order to arrive at the content that *she*, B, has Groat's disease, B needs to two things: Firstly, he needs to identify being the addressee of A with being an individual *to whom* A (*de re*) ascribes a property. This is a non-trivial step, since A only self-ascribes the prop-

²⁵A belief with a complex content, e.g. disjunctive, may have components that are not themselves believed, e.g. each disjunct separately, but in communication even an unembedded content may need to be grasped without belief, or any other attitude mode, attached to it.

erty of *having* a (unique) addressee with Groat's disease. If this step is allowed, B can advance as follows:

- (40) 6. B self-ascribes the property of being ascribed the property of having Groat's disease.

Exactly what would license this step separately is not obvious. Lewis's definition of what it is to ascribe a property to individual may be of help:

A subject ascribes property *X* to individual *Y* under description *Z* if and only if (1) the subject bears the relation *Z* uniquely to *Y*, and (2) the subject self-ascribes the property of bearing relation *Z* uniquely to something which has property *X* (1979, p. 539).

Applying this definition within her self-ascription 5., B may infer that, because of being the addressee of A, she also has the property of being ascribed having Groat's disease, i.e. 6..

Secondly, B needs to advance from the content of *having been ascribed Groat's disease* to the content of *having Groat's disease*. This second step cannot be taken without some kind of "disquotation", i.e. *being ascribed* Φ to *having* Φ . The Trusting step provides such a disquotation. The problem is that the Trusting step does not only take us from a content to another content: it takes us from a *believed* content to another *believed* content. And that is not what we want.

This comes out clearly if we let B self-ascribe an instance of a kind of disquotation principle:

- (41) B self-ascribes the property of: having Groat's disease if she is correctly ascribed having Groat's disease.

In order to isolate the content that *he*, B, *has Groat's disease*, B must detach by modus ponens, but this again means *accepting* the premise that the ascription is correct.

Clearly, nothing can give us what we want, since it is an inherent limitation of the centered content framework that (unembedded) contents cannot be self-predicated without being self-ascribed, i.e. believed. Thus, only in the world of the hyper-gullible, who accept a content even before knowing what it is, can centered contents with Weber's recentering offer a model of communication.²⁶

²⁶Weber suggests 2013, 212, n that the account could make do with something weaker than Trusting, such as "the speaker wants me to believe that she has made a true utterance". Clearly, the hearer can easily accept this, i.e. believe that the speaker wants her to believe that she has Groat's disease, without coming any closer to entertaining the content that *she* has Groat's disease.

For all this, it is in fact possible to account in a very indirect way, within the centered contents framework, and via the disquotation conditional (41), for the dialogue

- (39) A: You have Groat's disease.
 B: I don't believe that. / That's not true.

For, if B self-ascribes the property of *not* having Groat's disease, then by a property version of *modus tollens*, B can make an inference from (41), and self-ascribe the property of *not* having been correctly ascribed having Groat's disease. The inference would have the following format:

- (42) i) $\{\langle x, w \rangle : Fx \rightarrow Gx\}$
 ii) $\{\langle x, w \rangle : \neg Gx\}$
 iii) Hence, $\{\langle x, w \rangle : \neg Fx\}$

If the premises are self-ascribed, then the conclusion may be self-ascribed as well, provided other self-ascriptions are held constant. This is not exactly *modus tollens* as we know it, but it is certainly a valid inference in the sense that if any centered world $\langle x, w \rangle$ compatible with what B believes is such that $Fx \rightarrow Gx$ is true in w , and the same holds for $\neg Gx$, then it also holds for $\neg Fx$. Thus, together with the self-ascription 6., that B *has* been ascribed the property of having Groat's disease, she has what is needed for the second response in (39B): "That's not true".

Arriving at the basis for the other response, "I don't believe that", is yet more complicated. B must start from self-ascribing the property of *not* self-ascribing the property of having Groat's disease. Then B must self-ascribe the following property: if being ascribed the property of having Groat's disease, then (if not self-ascribing having Groat's disease, then not believing the ascription). Two applications of *modus ponens* yield the basis for the response.

Note that in both these responses, the hearer must already possess these self-ascriptions (of not having Groat's disease, and of not self-ascribing having Groat's disease, respectively), as well as the disquotation conditional (41), *not* form them as a result of considering the mere content, and also appropriately activate them in order to use them as premises in reasoning. It is unclear how that is done.

Note further that this way of accommodating the (39) dialogue presupposes that we do accept part of what Weber suggests, viz. that interpretation by the hearer is *meta-linguistic* and *meta-doxastic*; the hearer does not get the content from the speaker, but makes an ascription from a neutral point of view, of what the speaker self-ascribes. After that, a number of meta-doxastic inference steps must be performed in order in order to

arrive at the self-ascriptions that are the basis for the responses in (39B).

Not only is it very implausible that interpretation *ever* works like this, but since hearers who do not e.g. have the concept of self-ascription are able to perform feats as that in (39B) (at least the second), it must at least very often work in some other, simpler, way. To put it cautiously, the framework of centered-worlds contents does not appear to be able to account for this.

6. A mini-sketch of positive views

I shall here briefly sketch my own positive view. It is a Fregean theory, in line with the opening quote from Frege.²⁷

I implement this view in an extended possible-worlds semantics, which adds *doxastically possible* worlds—henceforth *D-worlds*—to the metaphysically possible worlds. That is, there are doxastically possible worlds that are not metaphysically possible—not *M-worlds*—e.g. a world where Heimson, as first-person presented to himself, is identical to Hume.

The crucial feature of the theory is that D-worlds are not all equally accessible. D-worlds where Lingens, as presented to himself, is anyone else than Lingens, are accessible only to Lingens. Analogously for Lauben. Metaphysically possible worlds are doxastically accessible to all subjects.

Lingens's self-presentation is then treated as an *intension* over the D-worlds accessible to Lingens, i.e. a function L such that $L(w)$ is an individual in w . In any M-world w , $L(w) = \text{Lingens}$, but as it is possible for Lingens to believe that he, as presented to himself, is someone else, in some impossible D-worlds accessible to Lingens, the value of L is someone other than Lingens.

The content of Lingens's first-person thought that *he* is Lingens is then a subset of the set of D-worlds accessible to Lingens, true at the M-worlds but false at some non-M-worlds. The content of *Lauben's* first-person thought that *he* is Lingens, is a subset of the set of D-worlds accessible to Lauben, false at the M-worlds but true at some others.

This Fregean view straightforwardly accounts for the possibility of *de se* thoughts. Self-locating thoughts, or *de se* thoughts, are distinguished by having special contents. Lingens's first-person thought that *he is Lingens* has a different content from his thought that *Lingens is Lingens*, and also from Lauben's first-person that that *he is Lingens*.

Lingens's first-person thought that *he is Lingens* is not available for Lauben to entertain. Similarly, my first-person thought that I am hungry is not available for you to think. The present view is therefore a variety of the *limited access* view that was rejected

²⁷My views here are similar to those of Jason Stanley (Stanley 2011, Chapt. 3).

by Perry. The question that naturally arises is whether this view isn't equally damaging to the possibility of communication as the proposals I have discussed above.

On my theory of communicative success, the answer is no. To present the view briefly by means of an example, suppose that John utters '*Ft*' to Mary. The *unstructured* meaning of '*Ft*', as uttered by John, is $\llbracket Ft \rrbracket_J$. This is also the content of John's *thought* that *Ft*. The *structured* meaning of '*Ft*' as uttered by John is

$$\langle \llbracket F \rrbracket_J, \llbracket t \rrbracket_J \rangle.$$

This is also the pair of corresponding conceptual elements in John's thought.

The corresponding structured content that results in Mary's mind is

$$\langle \llbracket F \rrbracket_M, \llbracket t \rrbracket_M \rangle.$$

The idea is that communication succeeds iff $\llbracket F \rrbracket_J \approx \llbracket F \rrbracket_M$ and $\llbracket t \rrbracket_J \approx \llbracket t \rrbracket_M$. That is, the meanings associated by John and Mary with the predicate *F* must be approximately the same, but need not be identical, and similarly for the singular term.

The similarity relation is transitive. On the present account, similarity amounts to identity within an antecedently fixed restricted subset of possible worlds. Call it the set of *normal* worlds.²⁸ Exactly how the set of normal worlds is fixed is a further question that would lead too far to discuss in the present context (cf. Pagin 2006). For present purposes, it suffices to say that only M-worlds, metaphysically possible worlds, are normal.

Thus, if John says to Mary

(43) I am hungry.

then communication succeeds just in case John's conception of himself and Mary's conception of John pick out the same individual (i.e. John) in all normal worlds, and John's conception of hunger has the same extension as Mary's in all normal worlds (and John and Mary combine these conceptions in the appropriate manner). Their respective thought content will indeed be different, but they will be similar enough.

I also claim that the same success conditions apply to the particular case of false identity statements like

(44) I am Hume

as said by Heimson (Lewis 1979, pp. 524-6). As an unstructured possible-worlds propo-

²⁸For a complete account, we should also restrict space-time to a relevant portion.

sition, it is the empty set within the domain of M-worlds, but the structured counterpart gives the hearer information about how it comes that the corresponding set is precisely the empty set. The hearer will not grasp how Heimson thinks of himself, but this is not needed. However, a fuller understanding of what it is for Heimson to have such a belief requires more than grasp of the structured proposition. An account of *this* understanding requires an account of the nature of the corresponding belief reports.

The reason this idea of communication is not just *ad hoc* is that it is needed for a reasonable account of communicative success independently of considerations of *de se* contents. To exemplify with general terms, it is well known that speakers differ in judgments of categorization (cf. for instance Hampton, Estes, and Simmons 2007).²⁹ The general idea is that conceptual differences do not matter to communicative success if they are irrelevant to the topic. John and Mary might have slightly different conceptions of a tree, and thus would differ in categorizing some possible entities, but that difference does not matter to the success of John's reporting

(45) Bill has two trees in his garden.

John's and Mary's conceptions of trees may differ, but they are guaranteed to agree on such mundane examples as can be found in people's gardens.

When it comes to singular terms, there is a discussion since Evans (1982, Chapt. 9) on the question whether sameness of *sense* (attached to the term by speaker and hearer) is necessary for communicative success, something that Evans, among others, denied. All in all, the idea that sameness of content between speaker and hearer is too strong a necessary condition for communicative success is not new, and is well motivated by considerations outside *de se* communication. Appealing to it is not *ad hoc*.³⁰

References

Beaney, Michael, ed. (1997). *The Frege Reader*. Blackwell Publishing.

²⁹As anecdotal evidence, according to an early-July 2012 Wikipedia article on *Horsetail*, giant Horsetail (*Equisetopsida*) in prehistoric times (about 350 million years ago) was a tall tree. However, according an early-July 2012 Wikipedia article on *Tree*, giant Horsetail was *not* a tree, because it did not have woody tissue. Clearly, the authors of these two articles did not have exactly the same conception of a *tree* (assuming they were both well informed).

³⁰Recanati (this volume, yy) expresses a similar view within his *mental files* framework: "The speaker's thought, like the hearer's, may be as indexical and centered as one may wish. The only thing that matters is that the speaker's thought, indexical and centered though it might be, determines in context the same objective truth-conditions as the hearer's thought formed upon understanding the utterance." Recanati, however, also combines this with a number of other views, some tied and some not tied to the mental files framework, and as a consequence, the overall accounts are rather different and even to some extent hard to compare.

- Egan, Andy (2007). 'Epistemic Modals, Relativism, and Assertion'. In: *Philosophical Studies* 133, pp. 1–22.
- Evans, Gareth (1982). *The Varieties of Reference*. Oxford. Edited by John McDowell.: Clarendon Press.
- Feigl, Herbert and Wilfrid Sellars, eds. (1949). *Readings in Philosophical Analysis*. New York: Appleton-Century Croft.
- Frege, Gottlob (1892). 'Über Sinn und Bedeutung'. In: *Zeitschrift für Philosophie und Philosophische Kritik* 100, 22–50. Translated by Herbert Feigl as 'On sense and nominatum', in Feigl and Sellars 1949, 85–102.
- (1918). 'Der Gedanke'. In: *Beiträge zur Philosophie des Deutschen Idealismus*. 100, pp. 25–50. Reprinted in Frege 1980. Translated and reprinted in Beaney 1997, pp. 325–45. Page references to the translation.
- (1980). *Funktion, Begriff, Bedeutung*. Göttingen: Vandenhoeck & Ruprecht.
- García-Carpintero, Manuel (2014 (forthcoming)). 'De Se Thought'. In: *Oxford Handbooks Online*. Oxford University Press.
- Gibbard, Allan (2012). *Meaning and Normativity*. Oxford: Oxford University Press.
- Hampton, James, Zachary Estes, and Sabrina Simmons (2007). 'Metamorphosis: Essence, Appearance, and Behavior in the Categorization of Natural Kinds'. In: *Memory & Cognition* 7, pp. 1785–1800.
- Kaplan, David (1989). 'Demonstratives'. In: *Themes from Kaplan*. Ed. by J Almog, J Perry, and H Wettstein. Oxford: Oxford University Press, pp. 481–563.
- Kölbel, Max (2013). 'The Conversational Role of Centered Contents'. In: *Inquiry* 56, pp. 97–121.
- Lewis, David (1979). 'Attitudes *de dicto* and *de se*'. In: *The Philosophical Review* 88, pp. 513–43. Reprinted in Lewis 1983b, 133–56.
- (1983a). 'Individuation by Acquaintance and by Stipulation'. In: *Philosophical Review* 92, pp. 3–32. Reprinted in Lewis 1999, pp. 373–402. Page references to the reprint.
- (1983b). *Philosophical Papers. Volume I*. Oxford: Oxford University Press.
- (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- (2004). 'What Experience Teaches'. In: *There is Something About Mary. Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Ed. by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar. Cambridge, MA: MIT Press. Chap. 5, pp. 77–103.
- Moss, Sarah (2012). 'Updating as Communication'. In: *Philosophy and Phenomenological Research* 85, pp. 225–48.

- Ninan, Dilip (2010). 'De Se Attitudes'. In: *Philosophy Compass* 5, pp. 551–67.
- Pagin, Peter (2006). 'Intersubjective externalism'. In: *What Determines Content? The Internalism/Externalism Dispute*. Ed. by Tomasz Marvan. Newcastle upon Tyne: Cambridge Scholar Press, pp. 39–54.
- Perry, John (1979). 'The Problem of the Essential Indexical'. In: *Noûs* 13, pp. 3–21.
- Quine, Willard Van Orman (1969). 'Propositional objects'. In: *Ontological Relativity and Other Essays*. New York: Columbia University Press, pp. 139–60.
- Recanati, François (2007). *Perspectival Thought. A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
- (this volume). 'Indexical Thought, Communication, and Mental Files'. In: *De Se Communication*. Ed. by Manuel García-Carpintero and Stephan Torre. Oxford: Oxford University Press, p. xx.
- Stalnaker, Robert (1978). 'Assertion'. In: *Pragmatics*. Ed. by Peter Cole. Vol. 9. Syntax and Semantics. New York: New York Academic Press, pp. 315–32. Reprinted in Stalnaker 1999, 78–95.
- (1981). 'Indexical Belief'. In: *Synthèse* 49, pp. 129–51.
- (1999). *Context and Content*. Oxford: Oxford University Press.
- Stanley, Jason (2011). *Know How*. Oxford: Oxford University Press.
- Torre, Stephan (2010). 'Centered Assertion'. In: *Philosophical Studies* 150, pp. 97–114.
- Weber, Clas (2013). 'Centered Communication'. In: *Philosophical Studies* 166, S205–23.