When we discussed the \textit{de se} character of indexicals, I said that \textit{de se} interpretations reflect a notion of self-location (Wechsler’s self-ascription). Today we consider the background for an approach to enriching the anaphoric account of indexicality to model their \textit{de se} character. In the relevant literature, \textit{de se} interpretations are often described as involving self-ascription or self-location. To give a feeling for why, we first briefly consider the semantics of locatives, including perspectival locatives, conventionally anchored to a salient point of view; and we consider how that notion is metaphorically extended to perspectival locatives in doxastic space.\footnote{See the Greek root δόκειν dokein ‘to appear’ ‘to seem’ ‘to accept’. Doxastic logic is the logic of belief.} Focusing on the doxastic perspectivals, we see that perspective can shift only in certain types of contexts, and that when it does, it always yields \textit{de se} interpretations.

We then review the way in which the \textit{de se} has been modeled by Lewis (1979) and Stalnaker (2008), using centered worlds. And we introduce a way of modeling perspectival anchoring in discourse, using a distinguished type of discourse referent, one associated with a salient doxastic perspective in discourse, which we’ll call a \textit{discourse center}. In the next lecture we’ll show how anchoring to discourse centers yields \textit{de se} interpretations for the doxastic perspectival expressions, including indexicals.

## 1. Locatives and perspectival locatives

We spent some time last week thinking about the semantics of demonstratives. The analysis I proposed takes a demonstrative to be a kind of locative, an expression that pertains to location. Locatives include prepositions like \textit{in}, \textit{at}, \textit{on}, \textit{under}, etc., and the locative prepositional phrases they can be used to form: \textit{in the room}, \textit{on the rug}, \textit{around the corner}, \textit{in front of the car}, etc., as well as adverbials like \textit{thereabouts}, \textit{catty-corner from the grocery store}, and, of course, \textit{here}. Locatives presuppose a three-dimensional model of space, and locate objects in that space with respect to reference points. In \textit{in the room}, the location of the relevant familiar room is the reference point; in \textit{catty-corner from the grocery store} it’s the location of the grocery store, etc. The simplest locative preposition is arguably \textit{at}, which just identifies the location with the reference point. Others, like \textit{on} presuppose that the object whose location is the reference point, e.g. the rug in \textit{on the rug}, has intrinsic properties: the rug has a top, and the reference point is adjacent to that top. \textit{catty-corner} presupposes that the object whose location is the reference point, the grocery store in my example, is located on a street corner, etc.
But some locatives have an even more complex semantics, presupposing more than just a reference point with certain properties: they are perspectival. A good example is behind the tree. Compare that to behind the car, where the reference point is the location of the relevant car. Like humans and houses, cars have an intrinsic back (and a front, and two sides); so we can take the reference location picked out by behind the car to be the space immediately adjacent to the back of the car (at whatever scale seems relevant); this intrinsic behind doesn’t depend on a point of view, but presupposes that the object has the property of having a backside. But a tree has no intrinsic front or back. So to understand what behind the tree means, one must adopt a point of view from which there is one surface of the tree that presents itself as front from that point of view, thereby determining what’s back. Then use of the PP presupposes such a point of view, as well as a trajectory from that point of view to the tree and beyond; the reference location the speaker intends is one that’s along that trajectory but on the other side of the tree from the observer standing at the point of view. Other perspectival locatives include to the left/right of. Some locatives can have either an intrinsic or a perspectival interpretation; behind the car is one: besides the intrinsic interpretation sketched above, it has a perspectival interpretation like that of behind the tree. But to the left/right and some others are dedicated perspectival locatives, their denotatum crucially a function of the presupposed point of view.

Demonstratives, as we characterized them last week, are dedicated perspectival locatives. Recall the semantics for this/that from Roberts (2002):

\[(39) \text{ Presupposed Content of Demonstrative } DP_1 \text{ (informal):}\]
\[
\text{Use of a (non-)proximal demonstrative } DP_1 \text{ presupposes that there is an accompanying demonstration } \delta \text{ whose unique demonstratum, correlated with a weakly familiar discourse referent } d_i \text{ by virtue of being demonstrated, lies in the direction indicated by the speaker at a (non-)proximal distance to the speaker.}\]

\[
\text{Semantic Content of Demonstrative } DP_i:\]
\[
\text{At any given world in the Common Ground, the semantic content of demonstrative } DP_i \text{ is the value assigned to } d_i, g(d_i).\]

Here, the point of view is the location of the speaker. In canonical use with an accompanying demonstration, the trajectory is given by that gesture. And the reference point is proximal/distal from the point of view along that trajectory. If the demonstrative includes a description, that description provides an additional, non-locative constraint on the denotatum, like the descriptive content presupposition of Heim (1982). The denotatum is then just the unique entity (satisfying that description) at the reference point along the trajectory from the point of view of the speaker.

In all the locatives just discussed, the space in question is physical space in the actual world. In textual or discourse deixis, the space is temporal—the two-dimensional space of utterance (corresponding to the arrow of time); the point of view is that of the speech time; and the trajectory is always the line established by the order of utterance from that “point of view”

---

2 This analysis is based on the work of Jefferson Barlew on locatives. See Barlew (2015), for his development of the formal semantics for locatives to model the notions of point of view and perspective in space.

3 Or at least, we don’t talk of them that way. Note that, e.g., trees in the northern hemisphere are asymmetric with respect to the position of the sun: moss grows on their north side, while the south side is far less likely to have moss on it. But we don’t conceive of this as lending trees the property of having a backside.
where the “future” is always immediately adjacent, as we commonly find with cataphora, and so only referred to with this.

1.1 Deictic motion verbs: Barlew’s (2017) analysis

But there are even more complex locatives, as reflected in their presuppositional content. One fascinating sub-class is that of the deictic motion verbs, like English come. Barlew (2017) studies come and a near-synonymous verb in the Bantu language Bulu, zu ‘come’. He shows quite clearly that come and zu are alike in presupposing that the reference point—the location of the intended destination of a movement from one location to another—is where the relevant agent believes (or fears, or hopes, or imagines... ) herself to be, and not necessarily where she actually is. Just to give the flavor of this, here’s just a bit of the relevant data from English.

“In [(75)]. . .the speaker and addressee are located at the destination of Guy's motion path, and the use of come is acceptable. [(76)]. . .shows that in a minimally different context in which they are not located at the destination of the motion path, the use of come is not acceptable.”

(75) [Context: Abondo (speaker) and Bella (addressee) are in Avebe (destination). Guy is in Ebolowa (source). Guy leaves Ebolowa, traveling to Avebe. As he is traveling, Abondo tells Bella:]
Guy is coming to Avebe.

(76) [Context: Abondo (speaker) and Bella (addressee) are in Avebe. Guy is in Ebolowa (source). Guy leaves Ebolowa, traveling to Kribi (destination). As he is traveling, Abondo tells Bella:]
#Guy is coming to Kribi.

But Barlew argues that, pace all the proceeding literature, what really matters is where the relevant agent takes himself to be, based on examples like the following:

(77) [Context: Last week, Chicago baseball player Ernie Banks was hit on the head. He is now a lucid amnesiac. After the accident, Ernie was transported to Boston to work with an amnesia specialist. For all he knows, he has never been to Chicago. He has been reading about the baseball player Ernie Banks, but does not realize that he is reading about himself. He reads that President Obama was in Chicago 3 weeks ago and met Ernie Banks. The doctor later tells her friend:]
a. #Ernie believes that President Obama came to Chicago.
b. Ernie believes that President Obama traveled to Chicago.

(78) [Context: Identical to (77), except that Ernie regains his memory.] Ernie believes that President Obama came to Chicago.

“In (77), the anchor, Ernie Banks, believes of Ernie Banks that he is located at the destination of the motion event. Nevertheless, (77a) is unacceptable because Ernie does not realize that he
believes this about himself. His belief is not \textit{de se}. In the minimally different context in (78) where Ernie believes \textit{de se} that he was in Chicago, the example is acceptable."

Barlew (pp.31ff, especially 38-43) argues that the possible anchors for \textit{come} are those with salient doxastic perspectives; this claim is backed up by corpus evidence from The Corpus of Contemporary American English, COCA, Davies 2008. It seems like the following types of context serve to make a perspective sufficiently salient that it can serve to anchor a deictic motion verb like \textit{come}:

- In discourse, the perspectives of speaker and addressee(s) are always salient. See (75) above and (79) below.
- In the complement of an attitude predicate the perspective of the agent of the attitude (usually expressed by the embedding subject) is salient. This was first noticed by Fillmore (1975) and Hockett (1990). Oshima (2006b,c) calls this “deictic perspective shift”. See (80) and (81) below.
- FID (Free Indirect Discourse). See (82) and (81) below.

Addressee’s perspective:

(79) [Context: Ann is in Cleveland, and Beth is in New York. They are talking on the phone. Ann says Where is John these days? Beth answers:] John is in Chicago. However, he is coming to Cleveland tomorrow.

Perspective of the agent of an embedding attitude:

(80) [Context: Ann is in Cleveland, Ben is in New York, and Chris is in Denver. On the phone, Ann asks Ben Where is John these days? Ben says:] Chris \{thinks/says\} that John is coming to Denver today.

(81) [Context: The interlocutors are in Columbus.]
   a. Ron says he's in New York, and he says Tom is coming to New York.
   b. Ron thinks he's in New York, and he thinks Tom is coming to New York.
   c. Ron is imagining he's in New York, and he's imagining Tom is coming to New York.
   d. Ron wishes he was in New York, and he wishes Tom was coming to New York.
   e. Ron is pretending he's in New York, and he's pretending Tom is coming to New York.
   f. Ron is fears he's in New York, and he fears Tom is coming to New York.
   g. Ron is wonders if he's in New York, and wonders if Tom is coming to New York.
   h. Ron hopes to be in New York, and he hopes Tom will come to New York.
   i. Ron is dreaming he's in New York, and he's dreaming Tom is coming to New York.

According to Barlew: “The key feature in these examples [in (81)] is that the anchoring implication is entailed by the information state relative to which \textit{come} is interpreted.”

He notes (2017:40): “Importantly, other perspectival expressions shift under attitude predicates in exactly the same way (Mitchell, 1986; Sells, 1987; Speas and Tenny, 2003; Oshima, 2006c;
Smith, 2009; Roberts, 2014), again providing support for the notion that come is a prototypical perspectival expression.”

FID facilitating deictic perspective shift:
(82) But with Mr. Ramsay bearing down on her, she could do nothing. Every time he approached - he was walking up and down the terrace - ruin approached, chaos approached. She could not paint . . . . . She rejected one brush; she chose another. When would those children come? When would they all be off? she fidgeted . . .

[from Virginia Woolf's To the Lighthouse, as quoted in Doron (1991:52), where Woolf is describing the inner state of her character Lily Briscoe]

(83) Meanwhile, Dick had himself just made his perpetual vows and was at home, preparing himself for the trip to France. (He had spent the preceding summer at St. Joseph's Oratory in Montreal, where he had made a start on French. He delighted in the pastoral work involved in helping the many pilgrims who come there in the summer.) He heard the news of my coming home, of my confused state of mind, and he was a little shaken. It took him only a moment to adjust, however, for he wrote that night to Father George S. DePrizio that his own desires remained unchanged. “I am anxious to study theology. . . ."

[from Novak, Michael. (2008) “The day my brother was murdered”. American Spectator 41:10, pp. 30-45 (COCA)]

About (83), Barlew says “Following Smith (2009), many expressions in (83) allude to Dick's perspective, raising its salience. These include the reflexive himself (see e.g. Sells 1987), the psych verb delighted, the perception verb heard, the emotional description shaken, the communication verb wrote, the discussion of Dick's own desires, and the direct quotation from his writing. Thus, expressions encoding information about Dick's perspective pervade the passage, making him an acceptable anchor for come.”

But Barlew argues that “non-interlocutors are not generally acceptable anchors for come” even when they are contextually highly salient. See (84), where despite the fact that the interlocutors are talking about Peyton Manning, he is not an acceptable anchor.”

(84) [Context: Ann and Beth are in New York. Beth says:] Sarah is in Chicago right now. However, she is a Denver Broncos fan and loves quarterback Peyton Manning who has a game there tomorrow. #Therefore, she is coming to Denver tomorrow.

Note that Peyton Manning the individual is salient; for example, he could felicitously serve as the antecedent for the pronoun he in a different final sentence: He always puts on a great show. Furthermore, Manning presumably self-ascribes being located in Denver. However, his perspective per se is not salient or relevant, and therefore he cannot be the anchor of come.
Barlew summarizes these observations about the anchoring of *come* as follows:⁴

(85) Anchoring implication (to be revised): According to some salient perspective of the anchor’s, the anchor self-ascribes being located at the destination.

Thus, “the perspectives relevant for the interpretation of *come* approximate the kinds of information states that a given anchor would have were she located at a certain place at a certain time.” (p.48)

Barlew then argues that the anchoring implication of deictic motion verbs is presupposed, not proffered, using tests like those from Partee (1984), including the possibility of quantificational binding of the presupposed anchor and its location:

(86) [Context: Leslie is describing some women, each of whom has a child who has caused a lot of heartbreak but who is now trying to be a better family member.]

> Every woman was glad that her wayward child came to Christmas dinner.

He also argues that the perspectival presupposition tends to project, as gauged by the usual family of sentences tests for presupposition projection. I refer you to his section 2.5 for details of that argument.

Barlew then argues in detail that Bulu *zu* ‘come’ is also perspectival in the relevant sense.

Note that the inventory of types of contexts that permit a shift in perspectival anchor from the default—the speaker—to some salient third person echoes what we observed about shifting indexicals in English (limited to a non-personal indexicals, mostly in FID and in some attitude contexts) and in languages with shifting 1st and/or 2nd person indexicals.

Other types of expressions which show variable anchoring limited to these contexts: evidentials (cf. “interrogative flip”, anchoring to the embedding agent in those languages that permit evidentials under attitudes), epistemic modals, and Pottsian Conventional Implicature triggers. We’ll return to the evidence for this claim in lecture 6.

What do these deictic perspective shifting contexts have in common which would lead us to expect this constraint on possible anchors for doxastic locatives, and on all these other types of variably anchored expressions as well?

---

⁴ See his further discussion for revisions that won’t concern us here. Among other things, he notes cases where the perspective is merely adopted by an agent, instead of reflecting where s/he believes she is, another type originally due to Fillmore (1975), such as (i) (adapted from Fillmore 1975: 67):

(37) [Context: The speaker is not on the island. She says:]

> There on the uninhabited island, waves lap upon the shore. Beaches sit empty as they have for millennia.
> a. ??Occasionally, a loon stays on the island (to roost) too.
> b. Occasionally, a loon comes to the island to roost.

And similar cases where the speaker describes what she sees at a distant site via webcam (his (35)), or what she pictures occurring at some location (his (36)).
Recall that in (77) above, anchoring to the embedding agent Ernie Banks gives rise to a \textit{de se} implication that cannot be cancelled. This is the key: The contexts which potentially license a shift in doxastic anchor are those that give rise to a \textit{de se} interpretation. That suggests, in turn, that the indexicals themselves are doxastic locatives: they are about self-location in the space of belief, i.e. the space of possibilities.

We sketched above how we might model the semantics of locatives. But how do we extend this to model doxastic self-location and the \textit{de se}? Centered worlds were introduced in formal semantics to give us the tools to model perspectival location in this metaphorical sense: self-location in the space of possibilities. We now turn to briefly consider how we use these tools for that purpose.

2. Centered worlds

Recall the original example that we used to illustrate the \textit{de se}:

(52) [Context: The baseball player Ernie Banks gets hit on the head and develops total amnesia. He doesn’t know his name or remember anything about his past, though he is lucid. During his long recuperation, he reads in the newspapers about a baseball player named Ernie Banks, and becomes fascinated with the guy’s career. His social worker reports to a nurse:] Ernie Banks thinks \textbf{he} is one of the greatest shortstops of all time.

The problem of \textit{de se} interpretations is a problem about how to model belief. What’s at issue is not so much how the world is—we all agree with Banks that Banks is a great shortstop—as Banks’ sense of where he himself fits into the way things are. There’s a sense in which one’s beliefs amount to a characterization of what’s possible and what’s true. But one aspect of what one believes is where one is located in the scheme of things. So self-attribution is about self-location in the space of possibilities consistent with one’s beliefs, what Lewis (1979b) has in mind when he says that the \textit{de se} involves self-ascription of a “locational property….in a certain region of logical space”. Banks is confused about where he is in the scheme of things.

The literature devoted to this phenomenon (Casteneda 1966,1967,1968; Morgan 1970; Lewis 1979b; Richard 1983; Perry 1993; Maier 2009; Ninan 2010, etc.) has focused on the use of centered worlds to model self-location and \textit{de se} interpretation in attitude reports. Earlier literature on the semantics of attitude reports largely drew on Hintikka (1969).

A Hintikka-style (1969) account of the semantics of attitude reports characterizes an agent \textit{a}’s attitude in a given world at a given time via a modal accessibility relation relativized to \textit{a}. A doxastic accessibility relation $\text{Dox}$ takes an agent \textit{a}, a time \textit{t}, and a world \textit{w}, and yields a set of worlds (or world-time pairs), those in which every proposition that \textit{a} believes at \textit{t} in \textit{w} is true. Stalnaker (2008) calls the output set of worlds the agent’s \textbf{belief set}, terminology we’ll adopt.

Lewis (1979b) argued that (52) and other similar examples show that this relation is not yet adequate to capture what it is to hold a belief \textit{when that belief pertains to oneself}. The
proposition that Ernie Banks is one of the greatest shortstops of all time is a set of possible worlds. To say that this proposition is true in all the worlds compatible with what Banks believes fails to capture the difference between his believing this as a proposition about himself and believing it about someone else.

Quine first introduced the idea of a centered possible world, and Lewis (1979b) adopts a variant to address the problem of de se interpretation. Very briefly, Lewis argues (p.518) that one who has a propositional attitude “locates himself in a region of logical space...He has a belief about himself, namely that he inhabits one of the worlds” where his belief is true. This is self-ascription of “a locational property: it is the property that belongs to all and only the inhabitants of a certain region of logical space.” Such propositions divide up the populace: some inhabit worlds where the belief is true, others do not. So the believer “has a partial opinion as to who he is: he is one of this class, not one of that class.” Thus Lewis takes this to be a kind of property: the property of inhabiting such and such a type of world. 5 But this property δ is equivalent to a set of centered worlds: pairs of a world and a “designated space-time point therein”; these are those pairs <c,w> in which the designated space-time point c corresponds to (a time-slice of) an individual which has the corresponding property δ, and in which the world w is one of those inhabited by c in virtue of his belief. Then the objects of attitude predicates are such properties or sets of centered worlds. Ninan (2010) succinctly characterizes how this works:

A centered world <w₀; x₀> is compatible with what an agent x believes in a world w iff x thinks in w that she might be x₀ in w₀ (x’s beliefs do not exclude the possibility that she is x₀ in w₀). An agent x believes de se in w that she is F iff every <w₀; x₀> compatible with what she believes in w is such that x₀ is F in w₀. To believe a centered proposition (set of centered worlds) p is for every centered world compatible with what one believes to be contained in p.

This account has the ability to distinguish between de se and non-de se contents. Following Egan (2006, 107), we can say that non-de se contents are boring centered propositions, where a centered proposition p is boring just in case for any world w, and inhabitants x, y of w, <w; x> ∈ p iff <w; y> ∈ p. Since boring centered propositions do not distinguish between worldmates, they are essentially equivalent to possible worlds propositions. A de se content is an interesting, i.e. non-boring, centered proposition. The account captures the idea that de se thoughts are not reducible to non-de se ones, since it allows for pure de se ignorance, ignorance that cannot be resolved simply by learning more and more boring centered propositions. [Ninan 2010]

Think of this as a matter of perspective: It’s one thing to know that a woman one sees reflected in the glass of a shop window is being approached from behind by a suspicious-looking character. That’s distressing, and you want to warn her. It’s quite another to realize that that woman is you yourself, and that the threat is behind you. Then, being warned that you are so-

---

5 Lewis also argues that the de se examples ultimately show that the objects of the attitudes are not propositions, but properties. Self location may bear not just on which world the agent is in, but on where she is in that world, so that her self-locating properties are not equivalent to pairs of an individual and a world. But as we’ll see below, Stalnaker argues that this argument does not go through; according to him, where a given individual is at a given time is a fact about the world she is in. If she might be in place p or in p’, then there is one world in which she is in p and a distinct world where she is in p’.
located, you go into action in a different way. As the work of Perry (1973, 1993) makes clear, this is the difference between having a map of what’s where, and having a map with a little red arrow on it that says “You are here”. Insofar as rational action is a function of the information available to the agent, plus her own goals and priorities, *de se* information is crucial.

How does such information amount to a perspective? What’s the formal analogy? A point of view in actual space, modeled as a three-dimensional Cartesian coordinate structure, can be characterized as a point in that space and a vector for which the point serves as origin—a location plus an orientation (Barlew 2015). Together, the coordinate structure, origin and vector amount to a way of modelling an agent’s perspective in physical space. The vector suggests an accessibility relation from the vector’s origin to other points (“locations”) in the space—those perceptually accessible from the adopted point of view plus orientation. It is an old idea (Lyons 1977, and many other references) to extend the locative notion metaphorically to model a more abstract notion of point of view and perspective. For *de se* interpretation, we can metaphorically generalize this locative model in order to characterize an agent’s doxastic point of view—how they see the world as characterized in their belief state, yielding their doxastic perspective on a matter.

Locative space is a set of points in a 3-dimensional Cartesian coordinate structure.

A **locative origin** is a point in locative space.

A **locative point of view** consists of a locative origin \( O \) and a vector \( V \) through the locative space, anchored in that origin: \(<O,VO>\)

A **locative perspective** is the set of points perceptually accessible from a given locative point of view: \( \text{Percep}(<O,VO>) \)

In the doxastic case, the “space” is that of possibility, as usual in formal semantics modeled as the set of possible worlds. In that space, call the origin of a doxastic point of view a (doxastic) **center** in a world, in keeping with the literature on the *de se*. A doxastic center is an ordered pair consisting of a doxastic agent—the anchor, an entity capable of being in a belief state—and a time. Then the origin of a doxastic point of view is the ordered pair of a doxastic center and a possible world, the world in which the agent has her beliefs at that time; this is a **centered world**: \(<<a, t>, w>\) , \( a \) the agent, \( t \) the time, \( w \) the world. Parallel with Barlew’s locative point of view, a doxastic point of view is a doxastic origin (the base centered world) plus a doxastic relation \( \text{Dox} \), which captures the agent’s orientation in the space of possibilities. The center’s doxastic perspective from the doxastic point of view anchored in the origin \(<<a, t>, w>\) is the agent’s **belief set** at \( t \) in \( w \), the set of centered propositions they take to be true, \( \text{Dox}(<<a, t>, w>) \).

This is a specification of the information (be it true or false) to which the agent has access from that point of view, each proposition a set of centered worlds. Continuing the spatial metaphor, a doxastic point of view yields the other “points” in logical space—the other possible centered worlds—to which the center has doxastic access. These are the worlds consistent with all the centered propositions in the agent’s belief set: each of these “points” is itself a centered world, and the center in the world is the “you are here” indication. The center is always some entity that exists in the world and which the agent takes herself to be in that world, “where” she sees herself in that “way things might be” (to borrow Stalnaker’s characterization of a possible world). Thus, a doxastic origin serves as the base centered world in a belief relation.
Summarizing:

Doxastic space is the set of all possible worlds W. A doxastic center is an ordered pair consisting of a doxastic agent \( a \) (the doxastic anchor) and a time \( t: <a, t> \). A doxastic origin is an ordered pair of a doxastic center and a world: \( <<a, t>, w> \). A doxastic point of view is a doxastic origin and its associated doxastic relation \( \text{DOX} \). A doxastic perspective is the information accessible from a doxastic point of view, \( a \)’s belief set at \( t \) in \( w \): \( \text{DOX}(<<a, t>, w>) \).

The model so defined permits us to capture what Stalnaker (2014) calls “self-location in thought”. I will argue as well that it is useful in interpretation even when de se attitudes are not obviously at issue, in modeling the CHARACTER of indexicals in a more flexible, yet constrained account.

Stalnaker (2008, in press) argues for a modification of Lewis’ view. He has multiple reasons for this, but here we will focus on one, which he calls a problem of calibration, crucial to comparing cognitive states. Hintikka’s approach to propositional attitudes via modal accessibility relations, sketched briefly above, makes possible comparison of the content of the objects of such attitudes across times and across persons: Two individuals A and B (or one individual at two times) believe the same proposition \( p \) at \( t \) in \( w \) just in case both \( \text{DOX}(A, t, w) \subseteq p \) and \( \text{DOX}(B, t, w) \subseteq p \). Then:

The unreconstructed Hintikka-style models of cognitive states took calibration for granted, but ignored belief change and self-location. The Lewis centered-worlds models recognized self-location, but provided no resources for representing the relations between informational states across time and across persons, and so no resources for clarifying the dynamics of knowledge and belief, or the communication of information between different subjects. The general framework that I am promoting allows for calibration across time, and across different subjects, but it also recognizes that calibration is a nontrivial problem, and may not be well defined in all cases. [Stalnaker 2012]

If beliefs are sets of centered worlds, and two distinct agents’ beliefs involve sets with distinct centers, how can we compare what they believe? If the same agent at distinct times corresponds to two distinct individual-time pairs, two centers, how can we compare what that agent believes at different times?

Stalnaker points out that replacing worlds in the Hintikka approach with centered worlds (below) permits us to talk about just one doxastic accessibility relationship \( R \), the agent given by the center of its first argument. Hence, a belief-state involving Lewis’ centered worlds can be characterized thus:

a belief state is a pair consisting of a centered world and its Dox-related belief set:

- the base (centered) world: the determining centered world, an ordered pair consisting of (a) the center: a person whose beliefs are being represented and the time at which she has those beliefs, and (b) the possible world in which the center has those beliefs.
• the belief set: the determined set, a set of centered worlds of the same type as the base world. In each pair \(<c, w>\), the \(c\) represents what the base subject takes herself to be in \(w\), a world which, for all she believes, may be actual. The worlds in these pairs are those which would be accessible from the base world under Hintikka’s doxastic accessibility relation relativized to the base center.

Stalnaker’s modified theory of centered worlds is realized with a model \(<W, S, T, \geq, E, R>\) where:

1. \(W\) is a nonempty set of possible worlds
2. \(S\) is a set of subjects or believers [my doxastic agents]
3. \(T\) is a set of times
4. \(\geq\) is a binary transitive connected anti-symmetric relation on \(T\), a relation that determines a linear order of the times.
5. \(E\) is the set of centered worlds meeting the condition that the subject of the center exists in the world at the time of the center, where
   - A center is a pair, \(<A, t>\), where \(A \in S\) and \(t \in T\). Subjects may exist at some times at some worlds, and not at others.
   - A centered world is a pair \(<c, w>\), where \(c\) is a center and \(w \in W\).
6. \(R\) is a binary relation on \(E\) that is transitive, Euclidean and serial. \(R\) must also satisfy condition (*), below. To say that \(<<A, t>, w>\ R <<B, t^*>, w'>\) is to say that it is compatible with what \(A\) believes at time \(t\) in world \(w\) that she is in world \(w'\), that she is person \(B\), and that the time is time \(t^*\).

\(R\) in 6 is a doxastic accessibility relation (the sort of relation I call \(DOX\) throughout this paper), representing a subject’s beliefs at a time in a world. The requirements on \(R\) guarantee that the agent has access to what she believes and to what she does not. See Stalnaker (2012, Chapter 2) for very useful discussion of the properties of this model. It has the additional condition (*):

(*) For any centers, \(c^*, c'\) and \(c''\), and worlds \(w\) and \(w'\): if \(<c^*, w> \ R <c', w'>\) and \(<c^*, w> \ R <c'', w'>\), then \(c' = c''\).

Intuitively, (*) tells us that “ignorance or uncertainty about where one is in the world is always also ignorance or uncertainty about what world one is in” [2012:70]. Stalnaker tells us that (*) is the main respect in which this model differs formally from Lewis’, in which the same center was permitted to occur in two “places” in the same world. Lewis argued that this was necessary in order to properly model the structure of de se beliefs, which he argued to be inherently more fine-grained than could be captured with possible worlds. But because of (*), for Stalnaker the contents of a belief state “can be taken to be ordinary propositions—sets of uncentered possible worlds, even though the centers determined by a particular belief state may play a role in determining which proposition is denoted by a that-clause with indexical expressions in it” (2008:71). That is, the “distinctive self-locating character [of self-locating beliefs] will be a feature of the subject’s relation to that content, and not a feature of the content itself” [my emphasis].

Suppose agent A says to agent B I am hungry; they may both grasp the same content of A’s assertion, a set of possible worlds. But qua centers at the time of utterance they stand in different
relationships to it: A, as the center of the base world of her belief report, is in an identity
relationship with derived centers which share a common property of hunger in all the derived
worlds at the time of utterance; whereas in B’s belief state resulting from acceptance of the
assertion, the derived centers needn’t share that property. So, among other virtues, Stalnaker
claims that (*) permits us to:

- model the denotations of the complements of attitude predicates as simple propositions,
  and thus straightforwardly:
- compare the beliefs of different subjects (whether self-locating or “boring”), and of a
  single believer at different times, something not possible in the Lewis model (where two
  sets of centered worlds with distinct centers were not comparable)
- model the way assertions change the context, as in Stalnaker (1979)
- model the dynamics of belief for a single agent, using standard belief revision theory

As proposed by Ninan (2010), common belief can also be modeled in this framework by taking
the center of the base centered-world to be a group such that all believe the propositions ϕ that
are true in all the accessible worlds, all believe that all believe those propositions, all believe that
all believe…, etc. The group can be modeled as a sequence of individuals, the sequences
representing where the members of the group mutually locate themselves and each other in the
possible worlds compatible with their common beliefs. Common Ground, as in Stalnaker
(1979), is more common acceptance, rather than common belief, coinciding with belief only “in
naïve conversations”. But otherwise it behaves like common belief. Thus, the interlocutors’
Context Set (CS) is a certain type of purported belief set, as defined in terms of this model. This
is the notion of (purported) doxastic state we will assume henceforth.

Thus, the denotation of the complement of believe can itself be a simple proposition, say, a set of
possible worlds. believe denotes a complex relation which takes the proposition and the agent
denoted by its subject and yields a relation between centered worlds. The base world is the pair
of the agent (at the time of utterance) and the world of evaluation; the derived worlds are those in
which the complement proposition is true of the self-identified counterparts of the agent. The
ambiguity of interest then is captured by whether a pronoun in the complement clause that’s
coreferential with the subject has a denotation in the belief worlds which is the same as that of
the derived centers, i.e. whether its denotation is taken by the base world’s center to be his
counterpart in the centered worlds in the belief set. This is the case in the 2nd of the logical forms
for (48) given below, yielding the de se interpretation, but not in the first:

\[
[\text{Ernie Banks believes he is a great shortstop}]^w \\
\text{believe}(\langle eb, w \rangle) \subseteq \{\langle \text{©}, w' \rangle | \text{eb is a great shortstop in } w' \} \quad \text{[non de se]}\\
\text{believe}(\langle eb, w \rangle) \subseteq \{\langle \text{©}, w' \rangle | \text{© is a great shortstop in } w' \} \quad \text{[de se]}
\]

In both interpretations, \{w' \mid \exists \text{©}: \langle w', \text{©} \rangle \in \text{believe}(\langle w, eb \rangle)\} \subseteq \text{[he is a great shortstop]}^{w,g},
where g(he) is Ernie Banks. That is, we can compare Banks’ beliefs on the two interpretations.
Thus, one of the virtues of this model is that because the content of an attitude-complement
clause can be characterized as a simple proposition, a set of possible worlds, we can ignore
centered worlds except when the centers are relevant to capturing truth conditions. In
considering the relata of a centered world \langle<agent, t>, w\rangle, where self-identity is not at issue
(there is no coreferential pronoun in the complement clause) we can abstract away from the
centers in the belief set, presenting the agent A’s doxastic state as a simple set of worlds. But if
the passage involves an attitude where the complement clause contains an NP coreferential with
A, and pertains to A’s view of who he himself is, then we consider the relevant centered worlds.

As proposed by Ninan (2010), common belief can be modeled by taking the center of the base
centered-world to be a group such that all believe the propositions that are true in all the
accessible worlds, all believe that all believe those propositions, all believe that all believe…,
 etc. The group can be modeled as a sequence of individuals, the sequences representing where
the members of the group mutually locate themselves and each other in the possible worlds
compatible with their common beliefs. Common ground, as in Stalnaker (1979), is actually
common acceptance, rather than common belief, coinciding with belief only “in naïve
conversations.” But otherwise it behaves like common belief in Ninan’s sense. Thus, the CG
itself is a certain type of (purported) belief set, a perspective in the technical sense, centered on
the interlocutors.

3. **Discourse Centers as perspectival anchors**

We will use centered worlds to model the *de se* character of indexicals in the truth conditions of
utterance which contain them. But what is the character of the presupposition(s) triggered by an
indexical which serve to assure that the indexical will yield a centered-world *de se* interpretation
in the relevant contexts? And how can the set of potential anaphoric antecedents for an indexical
be constrained so that only the interlocutors, the agent(s) of an embedding attitude or the
intended origin of FID are possible antecedents? Here I will argue that use of a distinguished
type of discourse referent, what I will call a *discourse center*, will satisfy both these requirements
at once.

One thing on which all accounts of indexicals agree is that these expressions are context-
sensitive. The anaphoric account I propose agrees, but crucially starts from a different notion of
context than that assumed by Kaplan, a notion independently motivated by a wide variety of
linguistic phenomena. His Character of an indexical is a function whose argument is the context
in which it is uttered, by which he seems to mean the *concrete situation* in which the speech act
occurs, modeled in terms of certain distinguished parameters (like speaker, addressee, time,
place, etc.). But not everyone would agree that this is how we should characterize a context of
utterance. E.g., see Stalnaker (1979), who argues that the relevant notion of context for
understanding presupposition satisfaction is the Common Ground (CG), a set of propositions,
each of these a set of possible worlds. The intersection of the CG is the Context Set (CS), the set
of worlds compatible with all the information in the CG. No one assumes that the Kaplanian
notion of context is adequate for treating presupposition generally. And see Lewis (1979), who
models context as an organized body of information, shared by the interlocutors, that changes
dynamically throughout discourse, his *scoreboard for a language game*, formally modeled as a
tuple of different types of information. I have argued extensively that a characterization along the
lines Lewis proposes is more appropriate for a theory of context-sensitivity in natural language
(see, e.g., Roberts 1996/2012, 2012b, 2004, 2015). So I take the context of utterance to be the
discourse context, an organized body of shared information. This body of information includes
information about the interlocutors’ Common Ground, but other kinds of information as well, as
we’ll see below.\textsuperscript{6} This more abstract notion of context is crucial to the empirically superior account of indexicals I will sketch here.

For one thing, the kind of context I will assume is dynamic, in that it may change in some respects in the course of interpretation, in ways that are crucial to interpretation. Again, this is unlike Kaplan’s notion of context, or Stalnaker’s CG, which are given once and for all per utterance. One of the most striking features of theories of dynamic interpretation is the uses of assignment functions to encode the discourse referents familiar to the interlocutors. Karttunen (1976) pointed out that we keep track of information in discourse about a number of individuals other than those we take to be actual entities in the world: for example, hypothetical individuals, or instantiations of quantificational statements. And we may keep track of information about someone we believe to be real, not knowing whether this individual is the same as some other with whom we are familiar, perhaps later merging information about the two. To distinguish these bundles of information from actual entities in the world, he called the former discourse referents. Heim (1982) and Kamp (1981) gave technical expression to this notion by modeling discourse referents as constraints on the assignment functions interlocutors can use to interpret variables at a given point in discourse. Discourse referents are variables in the sense that their interpretation is given by an assignment function. We assign an "address" to each bundle of information we take to be about a single (actual or hypothetical) individual, the address being a particular variable, say $x_{28}$—or $d_{38}$, ‘discourse referent 38’, to distinguish discourse referents from quantifier-bound variables. Then we require that any assignment function we use to interpret utterances at that point in the discourse be such that whatever it assigns as the value of $d_{28}$ be an individual which satisfies all the information we have about that (actual or hypothetical) individual discourse referent. Hence, instead of assignment functions being arbitrarily chosen, they are used to encode the information about discourse referents, permitting a continuity of reference.

Across discourse, these discourse referents, in the technical form of the set of permissible assignment functions—those compatible with the information at a given point in the exchange—are managed dynamically to reflect the way that information changes: If we add more information about a particular discourse referent, the permissible assignment functions will be further constrained—those in the preceding permissible set which are incompatible with the new information will be dropped from the set of permissible assignment functions that results from acceptance of that information by the interlocutors. But a discourse referent may be introduced under the scope of a quantifier (e.g., representing the variable bound by the quantifier) or modal operator (e.g. by an indefinite under the scope of the modal). If such cases where the discourse referent is merely an instantiation of a generalization or, say, hypothetical, triggered by information under the scope of a modal or other operator, then once the scope of the relevant operator has been closed off, outside its scope, the discourse referent is no longer available for reference, and accordingly, the relevant constraints on assignment functions no longer hold. This is the key to interpretation of donkey pronouns in these frameworks, as in (87):

\begin{itemize}
\item [\textsuperscript{6}] I have often been asked by philosophers whether the notion of context I propose couldn’t be modeled as information in the CG. Well, yes, of course it could, but that misses the point. There are distinguished \textit{types of information} which play distinct roles in interpretation and the evolution of the context itself, and this cannot readily be captured by throwing it treating it all as just a set of propositions.
\end{itemize}
Every man that owns a donkey beats it.

The discourse referent whose introduction into the information is triggered by utterance of a donkey in (87) will be available only under the wider scope of the universally quantified subject; that is, its “lifespan” will be limited to the scope of every. Hence it can bind the pronoun it (often called the "donkey pronoun") in the predicate, but normally is not available for reference after that. For example, we cannot felicitously continue (87) with I thought it looked miserable yesterday.7

Since the proffered content of a pronoun is a variable, the value assigned will be constrained by any information associated with that variable by the interlocutors at that point in the discourse. Moreover, utterance of a pronoun presupposes that there is already a discourse referent corresponding to that variable in the information structure. This is what Heim (1982) called the familiarity presupposition associated with a pronoun. A speaker cannot just walk in out of the blue and say She wants a new car. To be felicitous, this would have to be the continuation of some conversation we were having about some particular female individual. This is in keeping with our fundamental intuition that a pronoun is anaphoric: It carries us back to information presupposed to be already in discourse, and its presupposed content puts constraints on where to look for this information, thus giving clues to what the speaker intends. In this sense, the true antecedent of a pronoun is the corresponding discourse referent, an element in the information structure of the discourse.

So one of the elements of the context of utterance, conceived of as a tuple of bodies of information, is a set of discourse referents, DR, which is updated dynamically. Like Heim, I take discourse referents to be indexed bodies of information ("files") associated with actual or hypothetical entities about which the interlocutors share assumptions. An entity needn’t be mentioned in order for there to be a corresponding “file” on that individual; common knowledge of its existence ("weak familiarity") suffices. Hence, there are always weakly familiar discourse referents for the interlocutors in a discourse, for the time of utterance, etc.

In this essay, I want to keep discussion of technical details to a minimum, in the interest of focusing on the basic conceptual structure of the phenomena of interest. Hence, I will abstract away from the dynamics of discourse update, which can be complex to specify.8 To model the discourse scoreboard, I adopt a simplified and modified version of the information structure of Roberts (1996/2012), which offers a static snapshot of the discourse scoreboard at a given point.

---

7 Note that the above does not assume that quantifiers are unselective, as was the case in Kamp (1981) and Heim (1982). This facet of these early theories of dynamic interpretation has been discredited as leading, in the general case, to the wrong interpretation, though there still may be a sense in which it is appropriate in the cases for which it was originally proposed in Lewis (1975). See (Kadmon 1987,1990; Chierchia 1995) and references cited therein, and the discussion of the proportion problem below. As Groenendijk & Stokhof (1990) and Chierchia (1995) illustrate, dynamic theories are perfectly compatible with more classical, selective binding by operators.

8 Moreover, I tend to agree with Stalnaker, who has argued consistently over the past twenty-some years that the notion that we want might be not a dynamic semantics (in which, e.g., the meaning of an expression is its context change potential), but a dynamic pragmatics, where expressions have their usual intensional content and the dynamics are just those of updating and downgrading the information that’s contextually available for interpretation at a give point.
in the interchange, but now enriched with a set of discourse referents. But I'll give a few hints on how one might model this in a dynamic theory like the well-known Discourse Representation Theory of Kamp (1981), Kamp & Reyle (1993), or the compositional DyCG of Martin (2013, to appear).

Here is the scoreboard I assume:

**Context of utterance in a discourse D**: \(<D_D, Q_U D_D, C_S_D, D_R D, \alpha_D>,\) consisting of

- **DG**: the interlocutors’ evident goals, their Domain Goals
- **QUD**: the set of questions currently under discussion
- **CS**: the interlocutors’ Context Set, the set of worlds compatible with their CG
- **DR**: the set of Discourse Referents (DRefs), a set of variables of type \(\langle s, e \rangle\)
- **\(\alpha\)**: the set of discourse centers, each the ordered pair of a DRef and a time: \(\langle d, t \rangle\).

I will have nothing to say here about the interlocutors’ Domain Goals or the QUD; with one exception below, I will ignore them. As for the CS, the reader familiar with DRT can think of this as the set of worlds in which there are truthful embeddings of the corresponding DRS: those assignments of values to DRefs in its universe which are compatible with the DRS conditions. That is, as in Kadmon (2001), I assume that a DRS is richer than just a record of the conversation in progress, and includes all the information in the interlocutors’ CG.

As is usual in dynamic theories of interpretation, a discourse referent (DRef) \(d \in D_R D\) is a variable which serves as the address, or file label for information in the CG about some “entity” which is familiar in the discourse (Karttunen 1976, Heim 1982; cf. the reference markers of Kamp 1981). For an entity to be familiar in this sense, it need only be the case that the CS entails its existence.9 Thus, \(D_R D\) might be understood as the representational reification of the existential entailments in CS. If something is entailed to exist, then we track the information we have about that entity by introducing a “file” with that information. Because we track not only entities which we take to exist, but also hypothetical or arbitrary individuals under discussion (under the scope of a modal or other quantificational operator), DRefs cannot be reduced to actual individuals in a model. Instead, a DRef (file) acts as an address for tracking the information associated with what the interlocutors take to be a single (real or hypothetical) entity. The elements of DR, then, are correlated with individuals in the real world only if introduced in a *realis* context. Like free variables in static logical forms, elements of \(D_R D\) are interpreted by assigning them values in the model. The admissible assignment functions over \(D_R D\) are not arbitrary (as in the usual Montague grammar model), but are only acceptable insofar as they reflect the shared information associated with the DRefs, as reflected in the CS at the time of interpretation: Heim’s File Cards, Kamp’s truthful embeddings, and the related assignments in Martin’s DyCG.

There is one difference between the classical conception of DRefs in Kamp (his “reference markers”) or Heim, and DRefs in the present account. Their DRefs are of semantic type \(e\), so that admissible assignment functions map them to individuals in the model. But here DRefs will

---

9 This is the weak familiarity of Roberts (2003), discussed in lecture 1. See Martin (2013, 2016) for a formal implementation of this idea in his DyCG.
be of type \(<s,e>\), i.e. they are mapped onto individual concepts (ICs). Thus, in a given world in the CS, the interlocutors’ familiar DRefs correspond to Fregean senses, not extensions. One reflex of this is that it enables us to model what interlocutors know about what others believe about the entities with which the interlocutors are familiar. For example, one of the things my sister Jane and I know about Santa Claus is that her son Jake believes that Santa exists; so in any given world in our joint CS, the DRef \(d\) corresponding to Santa Claus will denote an IC whose value in all the worlds in Jake’s counterfactual doxastic state (as we know it) lives in the North Pole in that world, drives a sleigh on Christmas eve, etc. Take \(\text{DoxCS}(\text{Jake})\) to be those worlds which reflect all the things that the interlocutors agree that Jake believes; hence, for all \(p\) s.t. the interlocutors believe that Jake believes \(p\), \(\text{DoxCS}(\text{Jake}) \subseteq p\). Then because the interlocutors agree that they disagree with Jake about the existence of Santa Claus:

\[
\text{DoxCS}(\text{Jake}) \cap \text{CS} = \emptyset
\]

Now consider the possible values of the DRefs in our discourse context corresponding to Jake and Santa. These values are all those given by assignments \(g\) of values for those DRefs s.t. these values are consistent with the information in the interlocutors’ CS (and hence, in a DRS reflecting that CS would make all the DRS conditions true). Since DRefs are of type \(<s,e>\), then these \(g\) are s.t. given the Santa DRef \(d\), for all worlds \(w \in \text{DoxCS}(\text{Jake})\), \(g(d)(w)\) is an individual who has all the properties that we agree that Jake believes Santa Claus has. Then if Jane and I disagree about whether Jake thinks there’s a Mrs. Claus, there are worlds \(w, w' \in \text{DoxCS}(\text{Jake})\) s.t. \(|\text{married}|^w(g(d)(w))\) is true while \(|\text{married}|^{w'}(g(d)(w'))\) is false. In this way, the information associated with DRefs is cross-referenced, as it were, with the information the interlocutors share about others’ beliefs.\(^{10}\)

Besides this virtue of the new type for DRefs, I note without extended discussion that it permits the resulting theory to reflect many of the desirable features of the account of Elbourne (2005,2008), as well as facilitating incorporation of Aloni’s (2001) perspective shifting operation, discussed and suitably revised in Roberts (2014). And it will permit appropriate DRefs to represent the doxastic centers which play a role in indexical anchoring, even when those agents themselves are irrealis, as in iterated belief contexts. Nonetheless, for simplicity when there is no reason to take the value of a DRef to vary across the worlds in the CS, I will talk at such times about “the entity” which is the value of that DRef—i.e. its value in all the worlds in CS.

\(^{10}\) But suppose that Jake sees a fellow named Ortcutt walking on the beach with his family and thinks he’s a sterling citizen. On another occasion Jake sees Ortcutt late at night in a dive bar, wearing a trench coat and with his face mostly hidden by a pulled-down fedora, and Jake thinks this fellow must be a spy. It seems that Jake simultaneously has two distinct perspectives on this fellow Ortcutt, which lead him to incompatible beliefs about what is, in fact, the same guy. Then what do we say about the value of \(d\), the discourse referent for Ortcutt, in the worlds in \(\text{DoxCS}(\text{Jake})\)? From Jake’s point of view, these are not the same guy; hence Jake doesn’t know that his beliefs are inconsistent. If we want perspectives to be consistent, we might want to say that this situation is one where Jake has two distinct perspectives on what is actually the same individual; see §6 in Roberts (2014). As Quine (1956) discussed, this reveals a number of thorny issues pertaining to the semantics of belief. But since nothing requires us here to make firm decisions about how many perspectives a doxastic agent might bring to bear on a single individual, or even about the consistency of an agent’s beliefs, I’ll ignore these issues in the remainder of this discussion.
Now for an important innovation: The scoreboard contains a set of discourse centers \( \mathbb{C}_D \), a dynamically changing indication of those familiar doxastic centers (agents at a time) whose doxastic perspective the interlocutors take to be relevant at that time in the discourse. A discourse center \( \mathbb{C} \in \mathbb{C}_D \) is a pair \( <d,t> \) representing the agent \( d \) of a contextually relevant doxastic attitude at a given time \( t \). Both elements of this pair are DRefs, since times—or events—are taken to be also familiar in the context, following Partee (1984). For simplicity here, we’ll assume that times are comparable across worlds, so under any assignment of values to DRefs, \( t \) will denote a constant function. Thus, a discourse center is the correlate in conversational information terms of a center in Lewis’ centered-worlds framework. To distinguish centers, they are doubly co-indexed to reflect both the agent and time; so \( \mathbb{C}_{i,j} = <d_i,t_j> \). Accordingly, if \( j \neq k \), \( \mathbb{C}_{i,j} \neq \mathbb{C}_{i,k} \), since the same agent may have different beliefs at different times.\(^\text{11}\)

The set of centers in a discourse \( D, \mathbb{C}_D \), is as follows:

\[
\mathbb{C}_D \subseteq \{<d_i,t_j> | d_i, t_j \in \mathbb{DR} \wedge d_i \text{ is a doxastic agent whose beliefs at } t_j \text{ are under discussion in } D\}.
\]

Given indices \( i, j, k, l, m \) for familiar DRefs \( \in \mathbb{DR} \):

- \( \mathbb{C}_D \) always includes a distinguished center \( \mathbb{C}_{i,j}^* \), corresponding to the speaker(s) \( d_i \) at the time of utterance \( t_j \), and another \( \mathbb{C}_{k,j}^@ \) corresponding to the addressee(s) at that time.
- additional centers \( \mathbb{C}_{k,m} \) may be introduced in conjunction with the interlocutors’ consideration of alternative doxastic states, triggered in a constrained fashion either by lexical semantics or in accordance with the conventions of discourse styles like FID. In particular:
  - dynamic compositional interpretation of an attitude predicate conventionally triggers the introduction of a discourse center corresponding to the agent of the attitude at the event time of the holding of the attitude; and
  - discourse conventions associated with FID trigger introduction of a center corresponding to the agent whose perspective at a given time is adopted by the author.

In addition, like other components of the language scoreboard, \( \mathbb{C}_D \) is updated as the speaker changes or when leaving the scope of a doxastic operator or FID.

In a normal, “non-defective” context, \( \mathbb{C}^* \) will take the same value for all the worlds in CS—i.e., the interlocutors will assume that they know who the speaker is, in some sense of knowing who. If they know that John Perry is speaking, then the discourse referent corresponding to the agent of \( \mathbb{C}^* \) will be the IC whose value in all worlds compatible with CS is John Perry, a constant function. (More on what happens when the identity of the speaker is unknown, in Roberts (2014:§6).) This center is updated whenever the speaker changes. Just as in Partee (1984), where in order to adequately model temporal semantics and pragmatics in discourse, the Reichenbachian Reference Time RT and the utterance time are updated dynamically over the course of interpretation, in this framework the distinguished center \( \mathbb{C}^* \) at any given time is updated to be \( <d,t> \), \( d \) the familiar discourse referent corresponding to the speaker at that time

\(^{11}\)When we don’t think it will lead to confusion, we may suppress the relativization to times in the interest of simplicity, and talk about a Center \( \mathbb{C} \); as a distinguished type of discourse referent \( d_i \), corresponding to a contextually relevant doxastic agent. But technically this individual is the anchor of the center.
and \( t \) the time of utterance. The speaker is always the principal discourse center on the scoreboard—intuitively, the individual whose purported beliefs (presupposing the CG) are being reported in an assertion, whose identity serves as index for \( I \), whose location serves as default location for \( \text{here} \), whose doxastic perspective is the default anchor for an epistemic modal, etc., and whose perspective on the entities under discussion is the default assumed for the interpretation of NPs, at least in non-intensional contexts. Take \( \text{DOX}(©)(w) \) for a given world \( w \) to be the set of (for this illustration, non-centered) worlds consistent with the beliefs in \( w \) of the agent of © (the value of that IC \( g(d_k) \) in \( w \)) at the time of ©. Then assuming the speaker is sincere and competent and that a sincere speaker’s purported beliefs include the propositions in the CG, for any CS-consistent world \( w \) her belief state in \( w \) at the time of utterance, \( \text{DOX}(©*)(w) \subseteq \text{CS} \).  

At certain junctures in utterance interpretation, under DR update/downdate (Kamp 1981, Heim 1982, Partee 1984, Martin 2013) other discourse centers besides ©* may become salient, relevant, and accessible in the technical sense of Kamp (1981). First, when the speaker changes, the value of ©* changes, and similarly with the addressee and ©@. In some cases additional centers are introduced as a function of the compositional update of the context in the course of interpretation of a single utterance: For example, following Heim (1992), use of believe introduces quantification over worlds in the doxastic state of the subject. Then under the scope of this predicate, that agent (at the relevant time and world) serves as a subordinate center, ©believe, whose life-span qua center is the scope of the operator itself. Generalizing, an attitude predicate or other expression whose meaning involves quantification over the worlds in a belief state (or, like wish, is founded on such a belief state), always subcategorizes for a corresponding doxastic agent; typically with an attitude predicate, this is the denotation of the predicate’s subject (though see the object of convince in Carol convinced Georgia to cut her hair). Then under the scope of the attitude predicate \( R \), there will be an accessible discourse center ©\( R \) corresponding to that doxastic agent. This is crucial for capturing de se interpretations and indexical shifting in languages like Amharic and Zazaki, as I will illustrate later.

A pragmatic counterpart of this semantic mechanism serves to license shifted uses of English indexicals like here and now in FID. In FID, typically the author uses a variety of markers to indicate that this style has been adopted, and hence that the perspective has been shifted (see Eckardt 2014); then a new ©FID is pragmatically introduced to the set of salient centers, the agent of ©FID corresponding to the literary character whose perspective is to be adopted.

The introduction of discourse centers is more restricted than the introduction of discourse referents themselves. The latter (I have argued) requires only weak familiarity. But update of the set ©D occurs only (a) when the interlocutors’ roles change, (b) when subordinate centers are introduced as part of conventional update/downdate, as when triggered by attitude predicates, or (c) as a reflection of FID. There are, arguably, other doxastic operators besides those in the lexical semantics of attitude predicates and in FID that may induce ©D update. For example, English epistemic must is by default anchored to the speaker (e.g. see DeRose 1991) or, in interrogatives, the addressee; but under the scope of a perspectival adverbial like according to

---

12 Recall that as with a Stalnakerian CG, the CS is always \( \cap CG \). Hence, as the CG grows the CS shrinks—the more propositions in the CG, the fewer worlds are compatible with all of them.

13 As well as for addressing the de re interpretation of NPs in attitude complements; see Roberts (2014).
John, the anchor is shifted in (88), as it is in the complement of thinks with John as agent in (89). However, for introduction of a new center, it does not suffice that the most salient individual under discussion is a doxastic agent, even where that individual’s doxastic state is explicitly mentioned, as we see in (90), where the speaker is still the understood anchor of must.

(88) According to John, it must be raining.
(89) I just asked John what he thought about the weather. He thinks it must be raining.
(90) I just asked John what he thought about the weather. It must be raining.

We saw a parallel constraint in the doxastically anchored interpretation of come above. Thus, even though the set of discourse centers may be a proper superset of the set of interlocutors, it is far more constrained than the set of weakly familiar anaphoric antecedents, any additions and changes given either by compositional semantics or conventions of use like those associated with FID.

Centers other than ©* are subordinate to ©* just because in ordinary conversation the main goal of discourse is for the interlocutors to share content. The CG is the central doxastic perspective (or purported doxastic perspective) for the exchange, what speakers may always reasonably presuppose. Other agents’ views are relevant only insofar as they contribute to the developing shared CG. This is why ©* has a special status in discourse—it is always relevant, as reflected in its special role in indexical anchoring: Hence, in languages like English, a non-speaker center is subordinate to ©* in the sense that even in subordinate doxastic contexts it is still the distinguished ©* which serves as origin for indexicals like I. The addressee (as opposed to an eavesdropper) is also, similarly, a center ©© aware of his own role qua addressee. As they change roles, the set of discourse referents is updated accordingly to reflect the change.

As in standard theories of dynamic interpretation, we need to guarantee a tight relationship between CS and DR, such that the interlocutors track the information they share about the elements of DR, and use that information to constrain the interpretation of anaphoric NPs. Again, in the interest of simplifying exposition, in what follows I will assume the notion of a CS-consistent assignment g; such a function assigns to discourse referents in DR values which are consistent with the interlocutors’ information about them as reflected in CS. As in DRT or Heim’s File Change Semantics, tracking information about familiar discourse referents is accomplished by putting constraints on assignments of values to variables. The index associated with a discourse referent serves as its address, and as we add information about that discourse referent, this successively constrains the values which can be assigned to it in interpretation, requiring that all contextually felicitous assignments only give values which respect that information. Hence, if, say, a pronoun takes DRef d₁ as its antecedent, the value assigned to the pronoun by a CS-consistent assignment g for a given world w, g(d₁)(w), will be an entity which has all the properties the interlocutors take to hold of d₁ in w. This information is updated throughout interpretation, dynamically; and the lifespan of d₁ is restricted to the scope of any operators which have scope over the NP introducing d₁. Since all this information is in the interlocutors’ context for interpretation at a given point in discourse, we say that the felicitous assignments are CS-consistent.
One more point: If you are not familiar with work on dynamic interpretation, discourse anaphora, or *de se* interpretation, all this may seem like an exceptionally complicated elaboration of the Kaplanian notion of context, the simple situation of utterance. But what the careful study of these other phenomena have shown us over the past thirty years, is that the simpler notion of context is empirically inadequate. I like to say that

**anaphora is simple but discourse is complicated**

I.e., use of an anaphoric expression presupposes a discourse referent antecedent; its denotation, then, is a function of the contextually given value of that antecedent. But it becomes quite clear from the study of what factors constrain and facilitate anaphora and presupposition more generally that interlocutors in discourse track a complex body of data, updating it in the course of interpretation. And it is quite clear from the study of the *de se* that characterizing an agent’s belief state as a function from a world to a set of worlds is not adequate: we need (at least something like) centered worlds. So the only innovation introduced here is the dynamically updated set of discourse centers. But as I hope we’ll find time to discuss in the last lecture of this series, even those are independently motivated by the existence of a broad variety of types of perspectival expressions across languages, a great many of which presuppose a doxastic center in the sense defined above. So their use in the characterization of context of utterance is arguably independently motivated here as well.

**Additional references, not on course website:**

[http://dx.doi.org/10.3765/sp.9.5](http://dx.doi.org/10.3765/sp.9.5)