DEIXIS (EVEN WITHOUT POINTING)*

Una Stojnic
Center for Cognitive Science and Department of Philosophy, Rutgers
University

Matthew Stone
Center for Cognitive Science and Department of Computer Science,
Rutgers University

Ernie Lepore
Center for Cognitive Science and Department of Philosophy, Rutgers
University

1. Introduction

At a key moment in the "omelet" episode of Julie Child's *The French Chef*, we see a close-up of a hot skillet from above. A runny mixture of egg and butter sloshes around inside, while Child, working just beyond our field of view, shakes the skillet back and forth. As the eggs start to set, Child announces:

(1) There's your omelet, turning over on itself, forming itself in the bottom of the pan.

Child's use of (1) is an example of what we shall call a *situated utterance*. Situated utterances are used by speakers to comment on what's happening in a specific place and time, and to report on a specific perspective or body of information. In particular, to understand Child, we need to track the fact that she's describing the events presented in the accompanying video; we need to recognize that the newly-firm egg mixture we see *is* what Child means by 'your omelet' and that the shaking skillet we see *is* what Child means by 'the pan'.

This paper offers both a philosophical and a formal exploration of the interpretations of situated utterances. What exactly is the information conveyed by them? How much of their interpretation depends on linguistic knowledge and how much on extra-linguistic information? And which tools do we need to

formalize their interpretation and to distinguish the contributions of language and world knowledge?

We provide novel and surprising answers to these questions by appealing to the independently motivated theory of *coherence relations*. We argue that utterances can be related in qualitatively different ways to *situations* in the world, and that this relationship needs to be explicitly represented in the logical form of situated utterances. Doing so is necessary to capture the indirect and implicit contributions crucial to their interpretation. But it will also make it possible to explain, and to formalize, reference to real-world objects in situated utterances by invoking the same linguistic mechanisms that govern anaphoric reference in coherent discourse. Our formalism privileges discourse relations as a key point of contact between extra-linguistic information and grammar in situated utterances. We shall see that this assumption has powerful consequences for philosophical approaches to context, meaning and reference.

Our approach starts with the truism that coherent discourse is more than just a succession of unrelated sentences. In particular, new utterances in coherent discourse typically exhibit interpretive connections that link them to what's come before, and thereby, help to render the discourse as a whole sensible. To illustrate what we mean, consider an example from Hobbs (1979):

(2) John took the train from Paris to Istanbul. He has family there.
(3) John took the train from Paris to Istanbul. He likes spinach.

While a typical utterance of (2) is perfectly felicitous, a typical utterance of (3) is not. The contrast is straightforward: (2) does not simply state two random unrelated facts about John. It is a part of understanding of (2) that its second sentence provides an explanation of the information provided by its first; the hearer who fails to grasp this connection fails to fully understand (2). In the technical terminology of Coherence Theories (Asher and Lascarides, 2003; 2013; Kehler, 2002), a coherence relation of *Explanation* is part of (2)'s interpretation. This relation links the proposition expressed by the second sentence to that expressed by the first. Specifying this relation makes the content and organization of the discourse explicit: it indicates that the overall topic of the discourse is John's trip to Istanbul and that one purpose of the trip was to visit family. Crucially, understanding that the content conveyed by the second sentence provides an explanation of the content conveyed by the first one is part of successfully interpreting (2). Thus, inferring and understanding this connection allows the audience to understand (2) as a coherent whole, rather than just a string of unrelated sentences. The exact same expectation of coherence is present in (3); and *that's* why (3), out of the blue, seems infelicitous. Without further information, we are left wondering—is Istanbul famous for its spinach? Or Paris? Or does spinach cause a fear of flying? To fully understand (3), we need to figure out how its pieces fit into a coherent whole. We expect that the second sentence introduces an explanation—inferring the Explanation relation to hold—and are

left unsatisfied when we are unable to interpret the second sentence as providing one.

The idea we will pursue is that these same mechanisms of Coherence Theory can be exploited in devising a satisfactory account of the meaning representation of situated utterances, such as Child's (1). In all cases, we will suggest, interpretation includes identifying coherence relations that indicate how utterances fit together into a coherent whole. In situated utterances, coherence relations can tie utterances not only to previous bits of discourse but also, and crucially, to ongoing situations. In particular, we suggest that (1) gets its coherence, in part, as a report of what's visible in the situation captured in the accompanying video. We formalize this by including a corresponding coherence relation, which we call *Summary,* in the logical form of (1). This relationship again serves to render the content and organization of Child's discourse explicit. In particular, the sentence content will be related to the accompanying situation by the *Summary* relation only if Child's descriptions 'your omelet' and 'the pan' fit relevant objects that are visible there. In fact, what we see on the screen leaves no doubt about what's what. Thus, by including coherence relations, logical form captures many cases of reference in situated utterances, without positing special representations of the interpretation of referring expressions.

We develop this idea in more detail by exploring the interpretation of situated utterances with demonstrative terms, such as 'he', 'she' or 'that'. Consider the following ordinary case of demonstrative reference, with a speaker uttering (4), while pointing first at an object in the environment and then at the place where it should go.

(4) John put that there.[1]

The speaker's pointing action seems to be a crucial mediator of situated reference in such cases.

Traditionally, pointing is conceived of as a non-linguistic device that interlocutors exploit in pragmatic reasoning to supplement linguistic meaning. The claim is that for demonstrative terms, linguistic meaning alone does not determine a referent, whereas by contrast, for automatic indexicals, like 'I', it does. Uses of demonstrative terms need to be accompanied by additional, non-linguistic evidence that allows the audience to *identify* the referent; this evidence is conceived of as input to purely pragmatic processes of reference resolution, that are supposed to supplement the incomplete linguistic meaning. A pointing gesture can provide that evidence, but so can anything else that allows an audience to recognize the speaker's referential intention (Kaplan, 1989b; King, forthcoming a; forthcoming b; Neale, 2004; Reimer, 1992).[2] So, interpreting demonstrative terms involves open-ended reasoning about the speaker's mental states in the general case. We argue against this view in detail elsewhere (Stojnic, Stone & Lepore, 2013 ms.). In particular, we maintain that the reference resolution of demonstrative pronouns, on both anaphoric and deictic uses, is determined by

linguistic, grammatically encoded mechanisms that update and access an appropriately structured context. Moreover, we argue that in deictic utterances of demonstratives accompanied by the act of demonstration—the pointing gesture and its analogs—is itself a grammaticized constituent of the speaker's utterance that, together with the linguistic meaning (the character) of the demonstrative pronoun, determines the referent on an occasion of use.

An interesting remaining challenge for us is to explain which mechanisms govern reference resolution for demonstratives *unaccompanied by* overt demonstrative gestures; for an ordinary example, consider an utterance of (5), looking at the close-up of the stovetop from above:

(5) That's an omelet.

Even without an overt demonstrative gesture, (5) can still be successfully interpreted. Here is a case where traditionalists' appeal to open-ended processes of intention recognition might seem especially appealing, since the felicity of (5) seems to suggest that pointing is just one among many different avenues of available evidence exploitable in interpreting demonstratives, one among many different kinds of cues available for identifying a speaker's intentions.

We again disagree; even in these cases, we will argue, there are linguistic rules that determine the correct interpretation of a demonstrative. Key to our account, as we have already indicated in our introduction, are the interpretative connections—the coherence relations—that, as we shall argue, tie utterances not only to the ongoing discourse, but to real-world situations as well. In particular, because coherence relations make discourse organization explicit, they must be associated with operations that update the *attentional state*—that aspect of the context that specifies which entities are the most prominent, and thus, the most privileged as potential referents for pronouns and other demonstratives. For example, the fact that (5) is to be understood as a summary of what's happening in a particular situation directs attention to the most prominent entity in that situation as the best candidate for resolving the linguistic references in (5). Accordingly, if we formalize this connection using Coherence Theory, and model the demonstrative as referring to the most prominent entity in the context compatible with its meaning, we capture the correct resolution of the demonstrative in (5) transparently in logical form.

We proceed to develop our approach with three arguments in its favor. First, we will argue that our account is compatible with a range of intuitive cases where meaning seems to trump speaker intentions in interpreting demonstratives, something difficult to capture within an intentionalist framework. Second, we argue that our account is theoretically parsimonious: it involves straightforward extensions of independently motivated models of discourse. Finally, we argue that our approach captures meaning more precisely and provides a better account than its predecessors of what's needed to understand and disambiguate situated utterances.

2. Background and Foreshadowing

The contrast between our view and traditional views of demonstratives involves differing perspectives on the fundamental relationships among demonstrative reference, linguistic rules and speaker intentions. We begin by framing these differences more precisely.

On the traditional view, it is ultimately *the speaker's referential intention* that fixes the reference of a demonstrative in a context (see e.g. Kaplan, 1989b). This is particularly clear when there is no overt demonstration associated with a demonstrative usage, which is the case that most concerns us in this paper. In general, we call any view where intentions determine the reference of demonstratives *intentionalist*. On such views, the linguistic meaning of a demonstrative is incomplete: its does not determine a referent in context. A speaker's referential intention supplements linguistic meaning, and thereby, determines the referent of the demonstrative in a context.

Intentionalism offers an intuitively appealing way to describe demonstrative utterances and aligns closely with broader perspectives on meaning in philosophy. On intentionalist views, the incompleteness of the linguistic meaning of demonstratives explains the ambiguity and open-endedness we often encounter in interpreting them. Meanwhile, the view's appeal to speaker intentions makes reference resolution an essentially pragmatic process of intention recognition, and thus, explains how what intentionalists take to be non-linguistic demonstrative actions like pointing can serve to disambiguate demonstratives by providing evidence of speaker intentions. Finally, if speaker meaning is generally a matter of intention, as traditional views would have it (Grice, 1957), then it is not surprising that speaker intentions would determine meaning in a particular case.

Despite the appeal, it has proved deeply problematic to describe the contribution of linguistic meaning to the interpretation of demonstratives within an intentionalist framework. The problem arises most clearly with utterances where rules do seem to settle how demonstratives should be interpreted, and so, speaker intentions seem *not* to play the role intentionalism requires. For example, even if you intend Sue to be the referent of your use of 'she', if she is not a prominent candidate referent, you will fail to pick her out. For example, if you are pointing at Ann, then the referent of 'she' is Ann, *even if* you intend it to refer to Sue. And if you are pointing to Bill, then your utterance is infelicitous *even if* you intend 'she' to refer to Sue. Similarly, if you utter the sequence 'Mary walked in. She sat down', intending 'she' to refer to someone other than Mary, then, unless you do something to render this other referent prominent, the pronoun will anaphorically pick out Mary. Such cases clearly show that there are limits to what the speaker can reasonably intend the referent of a demonstrative expression to be on an occasion of use. This is constrained by rules governing the use of demonstrative expressions.[3] It is not clear why this should be if ultimately intentions determine their reference.

On one interpretation of these cases, (the one we think correct), intentionalists have gotten the relationship between speaker intentions and demonstrative reference exactly backwards. In choosing to use an utterance, a speaker must make rational commitments that anticipate how linguistic rules apply to that utterance. These commitments figure in the communicative intentions. When a speaker utters a sentence with a deictic occurrence of a demonstrative expression, she intends to convey certain information, and in particular, she intends the demonstrative expression to have a certain referent. But crucially she will have these intentions even if the linguistic rules alone determine the referent of the demonstrative! For, in that case, a rational speaker must work out which referent the rules specify, and commit to picking out that referent with the demonstrative to ensure that she is contributing the information she wants to convey with her utterance.

If a speaker intends to convey information about an object, say a, then she has to work out whether the rules in the given circumstance specify a as the referent of a demonstrative, and if so, by uttering the demonstrative, she'll commit to picking out a as its referent. If she wants to contribute information about a, she should use a demonstrative to do so only if the rules on that occasion specify a as the referent. If the rules specify something other than a as the referent of a demonstrative in the given circumstances, then the strategy of conveying the information about a by using a demonstrative in those circumstances will be ineffective. If that's right—and we will argue it is—then the mere fact that speakers have certain kinds of concomitant intentions when uttering certain linguistic expressions provides very little evidence about how demonstratives ultimately work. In other words, it is not controversial that a speaker normally has certain referential intentions when uttering a demonstrative. However, nothing about the ways in which the reference of a demonstrative is determined follows from this fact alone.

Of course, one can try to salvage intentionalism, as some authors have, by trying to accommodate the intuition that not anything goes, while preserving the gist of the account. This typically takes the form of placing constraints—inspired by broader ideas about how speaker meaning depends on intentions—on what a speaker can reasonably intend as the referent of an expression or on what kind of intentions count for determining the referent of a demonstrative expressions.[4] The general guiding idea is typically that by placing constraints on intentions, one can secure their public availability. Just as in the *Gricean* tradition, speaker meaning in general depends on publicly recognizable communicative intentions, so too the meaning of demonstratives depends on publicly recognizable referential intentions.

The trouble with this strategy is that to the extent that the public availability of an intention settles reference resolution, it seems that rather than the intention itself determining the referent, it is precisely the cues that make the intention public that are doing all the work. Indeed, the cues that make the intention publically available typically trump this intention when it comes to reference

determination; in fact, intentions are not even necessary. To see why, consider a speaker who intends to refer to Ann, but her hand becomes suddenly stuck, and so, she accidentally points at Sue, while uttering, "She is happy". It would be odd to say she intended to refer to Sue, or indeed, anything in the general direction of her pointing gesture. Quite clearly, though, intuitively, it is Sue, not Ann, who is the referent of 'she'. After all, the audience can follow up her utterance with "So, you are saying Sue is happy" and can challenge her with "That is false. Sue is not happy at all." (Note, the audience could not felicitously ask, "So, are you saying that Ann is happy?," or follow up with "That's false; Ann is not happy", or "True! Ann is happy".) The speaker cannot felicitously deny she said Sue is happy (or claim she said Ann is happy).[5]

In short, the content easily retrieved by the audience is that Ann is happy. If Sue were the referent, or the demonstrative failed to refer, why this is so would be a mystery. The moral is: even though a speaker typically has intentions to communicate information about a certain object, it does not follow that they *determine* the semantic content. And in fact, the examples suggest that these intentions are not even necessary for a demonstrative to successfully refer. Thus, we have no more reason to think intentions determine reference than that some other mechanisms do; and in fact, something even stronger is true: we have some reason to think that other mechanisms, rather than intentions, are what determines reference.

What, though, is a plausible alternative to intentionalism? If, in principle, any linguistic or extra-linguistic cue can supplement the meaning of a demonstrative in determining its referent on an occasion of use, that would be a reason to despair for anyone looking for a tractable semantic theory of demonstratives. Luckily, we suggest, a far more constrained account is preferable. We suggest that the reference of a demonstrative in a context is fully determined by the linguistic rules governing its use; in this regard, a demonstrative, like 'that', works much like a *Kaplanean* pure indexical, such as 'I'—its meaning is determined as a function of linguistic context. Fort this to work, however, the context in which the demonstrative is uttered has to be antecedently appropriately set up.

According to our view, the mechanisms that structure the context are themselves grammaticized linguistic mechanisms. Elsewhere we identified a set of such mechanisms that structure the context in a way that affects the interpretation of demonstrative expressions, in particular, when their occurrence is anaphoric, or deictic, but accompanied with a demonstrative gesture (Stojnic, Stone, and Lepore 2013 ms). Here, we want to suggest that the same story naturally extends to cases where the demonstrative is uttered deictically, *even when* no overt demonstrative gesture is present. Namely, the reference of a demonstrative is determined as a function of the context, where the context is antecedently set up by means of linguistic conventions.

In the remainder of the paper, we identify relevant rules, and offer an account of the relevant cases. Thus, although we do not here undertake the task of demonstratively refuting intentionalism, we suggest it's worthwhile exploring

an alternative: namely, the view that the referent of a demonstrative depends solely on the rules governing its use, quite independently of speaker referential intentions. If intentions were what ultimately determines the reference, it would be mysterious why in all the cases where the linguistic rules and intentions come apart, it seems that the rules trump the intentions. On our view, however, this fact is completely unsurprising, precisely because (even though intentions are typically present), it is the rules that determine reference. In the next section, we explore the conceptual and empirical underpinnings of this sort of approach.

3. A Coherence Account of Reference Resolution

We advocate an approach according to which demonstrative expressions, as a matter of language, receive their content in a context automatically (by an application of their character to the context at hand). For this idea to work, the context has to be antecedently set up correctly. One way this can be achieved is through the inferential connections that structure sequences of utterances into a coherent whole, as suggested by work on discourse coherence (Kehler, 2002; Asher and Lascarides, 2003).[6] The idea is that discourse is fundamentally comprised of relational contributions, which establish connections that link each utterance in the discourse by inference to segments of the preceding discourse. The interpretation of an utterance therefore implicitly refers to the interpretation of some prior discourse and comments on it. On coherence approaches, how an utterance attaches to a discourse determines which entities are prominent in interpreting it (Hobbs, 1979).

Kehler and his colleagues' (6) illustrates what's at stake (Kehler et al., 2008).

(6) Phil tickled Stanley, and Liz poked him.

If we analyze the second clause of (6) as a description of a *parallel* event to the one described by the first, we prefer to resolve 'him' to Stanley. But if we understand it to describe its *results*, then we prefer to resolve 'him' to Phil. For Coherence Theorists, these two interpretations of the second clause relate to the first, and it is a relation—*Parallel* in the first, and *Result* in the second— that suggests prominent resolutions for its references. Note that the discourse relations thus directly affect reference resolution. They also structure the discourse into higher-level units that shape possibilities for attaching subsequent utterances.[7]

Elsewhere, we have argued for capturing these effects as a matter of grammar. We achieve this by representing coherence relations at the level of logical form, as do Asher and Lascarides (2003) among others, and, crucially, further by enriching the formal specification of coherence relations to track the dynamics of attention in discourse (Stojnic, Stone and Lepore 2103 ms). Our idea is that once a coherence relation has been acknowledged, it is no longer optional what

the referent of the pronoun is—it is dictated by the rules of language, as part of the interpretive effect of the coherence relation. We'll briefly sketch here the principles required to develop this approach, because we build on these principles in our account of situated utterances.

First, we assume that context is structured to prioritize certain referential candidates over others. Following a range of research in discourse (notably Grosz and Sidner, 1986, and Grosz, Joshi and Weinstein, 1995), we refer to the ranking of referential candidates as the *attentional state* of the discourse. Our assumption is that the attentional state of the discourse—like other aspects of the context—changes over time. As new utterances are made and new information is contributed to the discourse, different candidates become prominent for reference. We formalize these changes in logical form. Thus, we will be able to read off of logical form what the ranking of prominent referents is that will govern the interpretation of any demonstrative expression.

Second, we assume that the attentional state of the discourse automatically determines the referents for demonstrative expressions (compare Grosz and Sidner, 1986 or Roberts, 2003). For example, we assume that the demonstrative pronoun 'him' automatically refers to the highest-ranked entity that satisfies its associated linguistic constraints: the referent must be a third-person singular male and its interpretation must be independent from certain syntactically superordinate expressions. Demonstratives can thus be formalized in logical form as functions from contexts to contents, so that their interpretation too can be read off of logical from. The interpretive rule for any use of the demonstrative pronoun 'him' corresponds to a function $f_{him}(c)$ that takes a context c with its ranking of prominent candidates and returns what this use of 'him' must automatically refer to in this context. (We relativize this to utterances of 'him' to track the third-person status of referents and the relevant inventory of syntactically superordinate expressions; the explanation is the same in the case of other demonstrative pronouns, modulo the difference in the associated linguistic constraints.)

Finally, we associate coherence relations with operations that update the attentional state. These operations are likewise visible in logical form. For example, with *Result,* the subject of the first sentence, describing the trigger, is placed at the center of attention and stays prominent throughout the description of the second sentence, describing the effect. This is what we see in the *Result* interpretation of (6), where 'him' refers to Phil. On the other hand, *Parallel* relations structure the context dynamically so that corresponding entities are prominent in corresponding positions across the two clauses. For (6), this puts Phil front and center as we consider the actions ascribed to the agent called 'Liz', and puts Stanley front and center as we consider the actions that befall the patient called 'him'. Again, this matches how we understand (6) on its *Parallel* interpretation.

In a sense, on our view, demonstratives are *not ambiguous*—discourse is. Moreover, ambiguities in discourse arise in tandem with questions about which primitive elements make up the discourse and how these elements are organized into higher-level units, on a pair with familiar lexical and syntactic ambiguities.

Thus, on our view, interpretation involves no further 'pragmatic ambiguities' to be resolved by intention recognition, once language users have resolved the grammatical ambiguities in a discourse and recovered its logical form. The process of recognizing a coherence relation just works as a process of disambiguation; viz., (6) is ambiguous between different discourses—each corresponding to a different interpretation—and a part of the interpretation process involves the relevant disambiguation.[8]

We propose to extend this framework to situated utterances by assuming that coherence relations connect utterances to the situations they describe. These situated coherence relations, like other coherence relations in discourse, are represented in logical form, and are associated with operations that update the attentional state of the discourse. In particular, situated coherence relations can focus attention on key entities in a described situation, just as other coherence relations can focus attention on key entities evoked by prior discourse. These entities then influence the interpretation of subsequent demonstratives. In this way, applying the general assumptions of our framework to situated utterances explains the possibility of deixis in these utterances without explicit pointing actions.

4. Coherence for Situated Utterances

To make this idea precise, we appeal to an expressive ontology of situations and eventualities (Hobbs, 1985; Kratzer, 2002). Situations are parts of the world, capturing particular states of particular objects, perhaps as located in particular spatial regions and changing over particular temporal intervals. Worlds are themselves a limiting case of a situation—the maximal (total) situation. We describe our ontology of situations and eventualities in more detail in the next subsection. We also need a formalism to characterize reference in discourse. We continue this section by explaining how we do this, by building on dynamic semantics, particularly the formalism of Muskens (1996). We conclude this section by exhibiting formal representations that capture the interpretation of situated utterances, including references to entities in the associated situations. These representations show precisely how our account resolves deixis according to linguistic rules. We close this section with a challenge to alternative approaches, by exploring the formal analysis of utterances where, we think, it's particularly attractive to capture interpretation through linguistic rules rather than through the intentions of the speaker or through coordination among interlocutors.

4.1 A Theory of Situations

Following Kratzer (2002), we adopt a logic in which situation variables are first-class citizens. This logic allows us to write expressions that not only characterize how the world is in general, but what's true in particular situations.

Informally, to say that a relationship is true in a situation is to say that the information in a particular situation is enough to guarantee that the relationship actually holds. This idea has a range of applications in formalizing common-sense judgments. For example, we can represent perception reports as relationships between individuals and the situations they observe (Barwise & Perry, 1983). We can represent counterfactuals as reporting on what the consequences would be if certain situations had not obtained (Kratzer, 1989). And we can use relationships among situations to capture the informational regularities that underwrite common-sense inferences, including both logical inferences and weaker, merely plausible inferences (Devlin, 1991).

Most importantly for our purposes, we can naturally appeal to an ontology of situations to capture the connection between situated utterances and those aspects of the world that they are used to comment on. For example, with (1), Child is not just giving the next step in making an omelet, or giving her audience new information about the principles of cooking. She's describing what's happening *in the situation* on the screen, in terms she expects her audience to confirm for themselves by examining what they see. An interpreter who doesn't recognize this about (1) has failed to understand it. Similarly, an utterance of (5) is a summary of a situation visible on the screen, and (consequently) the referent of its demonstrative is interpreted as the central entity in that situation.

Of course, the idea that some utterances can be used to describe particular situations is not new. It traces back at least to Austin (1950), who noted that some utterances do not seem to describe what's true in the world *tout court,* but rather, seem to make more fine-grained claims that characterize specific parts or aspects of the world. To see what's at stake, consider an example from Barwise & Etchemendy (1987). Two card games are taking place, one across town from the other. In the first game, Max is playing cards with Emily and Sophie, and in the second game, Claire is playing cards with Dana. Someone watching the first game mistakes Emily for Claire, and utters, 'Claire has the three of clubs'. In fact, in the second game, across town, Claire does have the three of clubs. So, what the speaker said is true of the world as a whole. Nevertheless, we might still judge that something has gone wrong with the speaker's utterance. Intuitively, the speaker was commenting on the card game he was watching and this card game is not correctly described with what the speaker said. In other words, intuitively, the speaker comment was about a particular situation—the one involving the game between Emily, Sophie and Max—and not about the world as a whole. Note, by the way, that the utterance 'Claire has the three of clubs' has a situated interpretation even though it lacks demonstrative noun phrases; it just involves a name and a definite description. The example underscores that we need an account of situated interpretation independent of any account of deixis.

Austin (1950) and Barwise & Etchemendy (1987) in turn assume that utterances are in fact about situations. We might call the situation that an utterance is about its *topic situation.* For an utterance to be accurate, the proposition the speaker expresses must be true in the topic situation. For example, in describing

the card game, the speaker's utterance is false because the proposition he expressed is false relative to the topic situation defined by the first game.

We build not only on the long philosophical tradition of analyzing utterance meaning as situated but particularly on the resources of Coherence Theory. In particular, Coherence Theory offers alternative, more flexible ways to relate utterances to situations. Some utterances get their coherence from their relationship to previous discourse, which allows the interpreters to understand how the utterances fit together, but we suggest, other specific kinds of utterances (and only these) get their coherence by linking up with what's happening in a particular situation and making a comment on it. In other words, we understand the speaker's comment in Barwise & Etchemendy's example as a comment on the specific card game not because all meaning is inherently situated (i.e. essentially about situations), but because the speaker uses this particular example with a particular *situated strategy* for achieving discourse coherence. As we pursue it, Coherence Theory is free to suppose that other utterances might not carry situated meaning. Moreover, Coherence Theory is free to separate the contributions of coherence relations from the truth conditions of an utterance—whether the utterance is situated or not. Coherence relations might just contribute updates to the attentional state, conventional implicatures, or other non-truth-conditional, backgrounded content. Nevertheless, Coherence Theory is committed to coherence relations being represented in logical form, because, importantly, *their effect is delivered by the grammar, constrained by linguistic rules, and interrelated with other aspects of meaning.*[9]

Our specific formal approach to situations follows Kratzer (2002) most closely. Kratzer offers a particularly appealing development of the links between situations, Davidsonian events, and the descriptions carried by linguistic expressions. Situations for Kratzer are *thin particulars.* That is, we don't have to think of situations concretely as capturing everything that happens in a particular region of space-time. Situations can be selective about which individuals they characterize and which properties and relations they specify about them. Accordingly, situations on Kratzer's theory are ordered by inclusion: this ordering respects the relationships among the space-time regions that the situations involve, but also orders situations based on the information that they make explicit. A situation s_1 is part of a situation s_2, only when s_2 contains all the information that s_1 offers.

Because situations have this general ordering, a key concept in Kratzer's ontology is that of a *minimal situation* that satisfies some proposition. Where propositions are sets of possible situations, a situation s is a minimal situation in which a proposition p is true if and only if it has no proper parts in which p is true. A situation s exemplifies a proposition p if and only if whenever there is a part of s in which p is not true, then s is a minimal situation in which p is true (Kratzer, 2002). Intuitively, a minimal situation has only what's necessary to guarantee that some proposition is true. A situation that exemplifies a proposition, by contrast, is in some sense fully described by a particular proposition, but may not realize that proposition in a minimal way.

In Kratzer's theory, eventualities, including Davidsonian events, turn out to be situations that exemplify propositions in characteristic ways. Bounded events, like "Joe ran a mile", correspond to minimal situations where the corresponding proposition is true; this fits the definition of exemplification immediately. On the other hand, unbounded events, like the event corresponding to "Joe is running", fit the definition of exemplification in a more general way. Events of Joe running aren't minimal situations of Joe running, but count as exemplifying situations because they have no subpart where "Joe is running" is not true. This suggestion respects the insight from aspectual semantics that events of Joe running involve homogeneous situations, all of whose parts also involve Joe running.

This ontology makes it particularly easy to characterize the relationship between events, situations and linguistic descriptions. Suppose we have some linguistic description. An event will be a situation that exemplifies this description. By contrast, the description will be true in a wide range of larger situations that specify this information as well as additional information. However, each of these larger situations will contain an exemplifying eventuality as a part. Semantically, then, the key link between a description and a situation that it is about is that the situation must contain a suitable eventuality that exemplifies the description as a part.

4.2 Dynamic Semantics for Discourse Anaphora

To capture reference across discourse, we build on Dynamic Semantics—a successor to approaches to discourse semantics inaugurated by Kamp and Heim in the early 1980s (Kamp, 1981, Heim, 1982). These approaches start from the observation that semantic theories need to track interpretive dependencies between sentences to correctly describe the truth conditions of discourse. For example, take (7):

(7) A man walked in. He sat down.

The most prominent reading of (7) links 'he' in the second sentence to 'a man' in the first. On this reading, (7) is true if and only if some man walked in and sat down. Dynamic semantics captures this interpretation in a compositional way by treating possible interpretive dependencies across discourse as part of an evolving context. Contributions to discourse are modeled formally as updates that can access and change this context.

The general theory of coherence in discourse requires a particularly expressive form of dynamic semantics. For example, Asher and Lascarides (2003) develop their theory using a formalism in the dynamic tradition called SDRT for Segmented Discourse Representation Theory. SDRT includes meta-level mechanisms for naming updates to the context, relating them to one another, and thereby, establishing the meaning and structure of discourse. The SDRT formalism could be extended straightforwardly to describe situated utterances in

much the same way we do here. However, our ontology of situations makes it possible to present the key ideas of our approach in a much simpler formalism. Thus, we shall pursue the latter, simpler strategy.

We follow Muskens (1996) and Dekker (1994), with extensions drawn from our own work on reference in discourse (Stojnic, Stone and Lepore, 2013 ms).[10] On this approach, contexts are modeled as sequences of entities,[11] ordered by their attentional prominence.[12] We'll use the typical variables for assignments, g, and h, to notate these sequences. By tracking the attentional state in this way, it becomes possible to define expressions that access entities as a function of their prominence in context. One way to do this is to pick out entities directly by their position in the ordering. We assume we have a variable 'x_i' for each position i that picks out the ith element of the attentional state: 'x_0' for the most prominent entity, 'x_1' for the second most prominent entity, and so forth. As we explained in Stojnic, Stone and Lepore (2013, ms.), we use these variables to specify the grammatically determined argument structure of predicates. In formal terms, the basic way of specifying information in our dynamic language is through conditions of the form $r(v_0 \ldots v_n)$, where 'r' is a relation symbol and 'v_0' through 'v_n' are variables that access the attentional state (i.e. the dynamic context). The condition $r(v_0 \ldots v_n)$ is true in context g if, and only if, the tuple $g(v_0) \ldots g(v_n)$ is an element of the interpretation of r.

Normally, the most prominent entity x_0 and the first argument to relation symbols will be a situation or eventuality argument. In keeping with Kratzer's ontology of situations, we understand the conditions $r(x_0, v_1 \ldots v_n)$ corresponding to natural language meanings to impose the constraint that x_0 contains an eventuality that exemplifies the underlying state of $v_1 \ldots v_n$ standing in relation r. This way, when we conjoin multiple conditions to capture the meaning of a complex sentence, we get a specification of a single overall situation encompassing all the information in the sentence. Meanwhile, because of the importance of situations and eventualities on our account, and also to make our formulas more readable, we will use 'e_i' as an alternative notation for 'x_i'. We will use the distinction as a mnemonic for whether an argument position involves a situation/eventuality or an ordinary individual. Thus, typical conditions will have the form $r(e_0, v_1 \ldots v_n)$.

In addition to these variables, which are used to represent argument structure, we also have expressions that correspond to the interpretation of demonstrative expressions. These expressions pick out entities from the attentional state based on the properties of those entities as well as by their place in the ranking. For this paper, the key case is an individual expression '**@that**', which we assume picks out the most prominent non-human ordinary individual in the attentional state. (This fits the meaning of 'that' in the examples we consider.) We design our logical forms to include operations raising entities to prominence in ways that ensure that **@that** is always resolved to its intuitively correct referent.

More precisely, then, the meanings of utterances take the form of updates that change the state of the discourse. An update is interpreted as a relation between an input context and possible output contexts that can be obtained

by incorporating the changes in information and attention that an utterance contributes. The simplest updates simply contribute information, as embodied by some condition *C*. Such updates are written [*C*]. Formally, these updates are interpreted as a partial identity relation. That is, [*C*] (*g,h*) if and only if $g = h$ and the condition *C* is true on the interpretation of the variables given by *g*.

We represent the contribution of indefinites and demonstratives in the same fashion as in earlier work (Stojnic, Stone, and Lepore, 2013 ms). Both indefinites and definites affect the prominence ranking of the entities in the attentional state, by introducing or promoting entities in the ranking. Indefinites update the attentional state by inserting an entity non-deterministically at a specified place in the ranking. Our notation for this update is $\langle \alpha k \rangle$. Intuitively, the effect of an indefinite update is to introduce a nonspecific entity at a specified position *k* in the attentional state, while all the entities that were in a position *k* or lower in the ranking prior to the update, are pushed one position lower in the ranking. Formally, $\langle \alpha k \rangle$ (*g,h*) if and only if $g(i) = h(i)$ for $i < k$, *h* has any entity at position *k*, and $g(i) = h(i+1)$ for $i \geq k$.

By contrast, the interpretation of demonstratives updates the attentional state by updating the ranking with a specific entity that's determined by the grammar. Our notation for such updates is $\langle \pi k c \rangle$, where '*c*' is an individual denoting term. Intuitively, the effect of a demonstrative update is the introduction of an entity that's specific—by contrast with the existential update associated with indefinites—at a specified position *k* in the ranking, while all the entities that were at a position *k* or lower in the ranking prior to the update are pushed one position lower in the ranking. Formally, we have $\langle \pi k c \rangle (g,h)$ if and only if $g(i) = h(i)$ for $i < k$, $h(k)$ is given by the interpretation of *c*, and $g(i) = h(i+1)$ for $i \geq k$. Obviously, (given that our formalism, for simplicity, does not distinguish between presupposition and assertion) $\langle \pi k c \rangle$ could be written by composing an indefinite update $\langle \alpha k \rangle$ with the condition [$x_k = c$]. We choose to define the two kinds of updates separately, however, to reflect the fact that qualitatively different mechanisms may be involved in indefinite and demonstrative reference in language. In particular, we will capture the effect of a pointing action designating a particular individual *c* by a suitable update $\langle \pi k c \rangle$ in logical form.

In addition to these tools, we add a further update $\langle \sigma k s \rangle$ to our language for this paper to capture reference that's mediated by individuals' prominent place in a situation *s*. Intuitively, $\langle \sigma k s \rangle$ updates the entities in the ranking beginning at position *k* to include the central entities in the situation *s* of certain key types—we need at least the central individual, the central location, and the central event—in set conventional order while all entities occurring at a position *k* or lower prior to the update are pushed three positions down the stack. More formally, letting *u(s)* be the central individual in *s*, *l(s)* be the central location of *s*, and *q(s)* be the central event in *s*, then $\langle \sigma k s \rangle (g,h)$ if and only if $g(i) = h(i)$ for $i < k$, $h(k) = u(s)$, $h(k+1) = l(s)$, $h(k+2) = q(s)$ and $g(i) = h(i+3)$ for $i \geq k$.

Finally, we note that in a familiar fashion two updates *H* and *K* can be sequenced into an update *H*; *K* that performs the update given by *H* followed by

the one given by K. So, formally $H; K (g,h)$ if and only if there is some j such that $H(g,j)$ and $K(j,h)$. Updates are interpreted through an implicit operation of existential generalization achieved via quantification over all possible assignment functions.[13] That is, a discourse is true in an input context if and only if the input context is related to some output context by the meaning of the discourse.

The last ingredient we need to complete our formal account is to spell out the effect of coherence relations in situated utterances. As suggested by our examples before, coherence relations affect what entities are prominent for subsequent reference and anaphora. The coherence relations that are operative in situated utterances are no exception to this. Thus, we assume that the coherence relation is represented in the logical form partly with a condition capturing the interpretative connection between the utterance and a specified situation, and partly by an update to the attentional state of the discourse capturing the change in focus inherent in commenting on a specific situation in this particular way. For example, we will commonly appeal to the combination $\langle \alpha 0 \rangle; [Summary(s_0, e_0)]; \langle \sigma 1 s_0 \rangle$ to express the interpretation of situated utterances like Child's (1) or its variant (5). The idea is that these utterances introduce and describe a situation e_0 that offers a summary of some accompanying situation s_0. At the same time, the coherence relation updates the attentional state to make central entities from s_0 prominent for resolving references in the speaker's characterization of e_0. More precisely, as with all coherence relations, the information represented by the predicate *Summary* reflects semantic and pragmatic constraints. Semantically, e_0 must be part of s_0. Following Kratzer (2002), this entails that the information in the accompanying sentence, which is fleshed out in terms of constraints on e_0, all winds up true in s_0. Pragmatically, *Summary*(s_0, e_0) holds only if the information the speaker uses to characterize e_0 provides a good answer about "what's happening" in s_0. The idea is that summary appeals to broad, basic categories to provide essential information. We have in mind something like the "vital nuggets of information" needed to answer definition questions (Voorhees, 2003).

Of course, not all situated utterances offer a *summary* of an unfolding situation. For example, utterances can also offer *assessments* that invite the audience not to define what's happening but to appraise it. Take "Yummy!" as an example. In commenting on the food in this way, the speaker expects the audience to join in her appreciation. That is a part of understanding the utterance. A formal characterization of *Assessment* would appeal to the semantics of predicates of personal taste and the distinctive pragmatic functions of such judgments, perhaps, following Crespo and Fernandez (2011). And speakers can also link up questions and instructions to ongoing activity by suitable relations.

4.3 Worked Examples

Putting all this together, then, we capture formally the desired interpretation of (4) by specifying the dynamics of discourse as in (8), using c_1 for the referent

of *that* and c_2 for the referent of *there*, where both occurrences are accompanied by a demonstrative gesture.

(4) John put that there.

(8) $\langle \alpha 0 \rangle$; $\langle \pi 1 j \rangle$; $\langle \pi 2 c_1 \rangle$; $\langle \pi 3 c_2 \rangle$; $[\text{put}(e_0, x_1, x_2, x_3)]$

(In (8) 'e_0' denotes the event of putting, 'j' denotes John, 'c_1' is what 'that' points to, 'c_2' is what 'there' points to.)

Things obviously become more interesting once we factor in coherence. For the utterance of "That's an omelet", we offer (9) as its formal representation, which defines the central entity in situation s_0 as an omelet:

(9) $\langle \alpha 0 \rangle$; $[\text{Summary}(s_0, e_0)]$; $\langle \sigma 1 s_0 \rangle$; $\langle \pi 1 @\textbf{that} \rangle$; $[\text{omelet}(e_0, x_1)]$

As explained in the previous section, the update $\langle \sigma 1 s_0 \rangle$ formalizes how the discourse relation renders entities prominent for reference, just as we observed in (6). The discourse relation *Summary(s_0, e_0)* captures the interpretive connection between the utterance describing e_0 and what's happening simultaneously on the screen in situation s_0. Such updates can capture the interpretation of demonstratives when there's no explicit pointing or demonstration in the utterance. Namely, it is the effect of a *Summary* that the central entity of a situation the summary is about is rendered most prominent; by the update $\langle \sigma 1 s_0 \rangle$ in (9), the omelet is added as the most prominent object in the attentional state. Thus, when we examine the context after this update to interpret $@\textbf{that}$, we are led to this omelet as its referent.

Meanwhile, we formalize the relevant interpretation of (1) as in (10).

1. There's your omelet, forming itself in the bottom of the pan.

(10) $\langle \alpha 0 \rangle$; $[\text{Summary}(s_0, e_0)]$; $\langle \sigma 1 s_0 \rangle$;

   $\langle \alpha 1 \rangle$; $[\text{omelet}(e_0, x_1)]$; $[\text{yours}(e_0, x_1)]$;

   $\langle \pi 2 @\textbf{there} \rangle$; $[\text{loc}(e_0, x_1, x_2)]$;

   $\langle \alpha 2 \rangle$; $[\text{pan}(e_0, x_2)]$; $[\text{forming-self-in}(e_0, x_1, x_2)]$;

Now when we interpret the deictic "there", we look in the current context for the most salient location; this is still the location that's contributed by the attention-shifting update $\langle \sigma 1 s_0 \rangle$ associated with the coherence relation: the place portrayed in the situation that the utterance is describing. Meanwhile, we interpret the noun phrases responsible for characterizing the omelet and the pan that Child's utterance describes in the ordinary way, but crucially the meaning of *Summary* requires that we find the omelet and the pan in the situation we see.[14]

As suggested above, combining situated representations with discourse relations, specifically as in (9), makes it possible to better characterize the logical form of demonstrative utterances in otherwise problematic cases. In particular,

it captures how speakers and interpreters can rely on *the world* to disambiguate what they say and to understand one another.

Here's a telling case. It's the beginning of spring, 2012, and Jupiter and Venus are shining brightly very close together—just a few degrees apart—in the evening sky. The speaker has deployed a telescope facing a window over the western sky. When a visitor arrives, the speaker adjusts the telescope, then says, without any further demonstration, either (11) or (12).

(11) That's Jupiter. You can even see four moons.
(12) That's Venus. You can see the crescent.

We (and our informants) find these utterances unproblematic. But Coherence Theory is required to get their interpretations right. These are comments on what's visible through the telescope. You can't see four of Jupiter's moons or the crescent of Venus, and distinguish the two planets apart with the naked eye, and the speaker isn't suggesting otherwise. Moreover, the coherence relation and its recognition is what makes it possible for the speaker to refer to Jupiter or Venus as *that*. Unless the audience recognizes that the speaker is making a comment about the view through the telescope, they will fail to understand the utterance; but more importantly, realizing that is all there is to the interpretation of (11) and (12). Given the astronomical conjunction, one cannot know, without looking through the telescope, which of the two objects the speaker is referring to; but once the audience recognizes that the speaker is making a summary of a scene viewed through the telescope, there is *no* further ambiguity of the pronoun to resolve. The referent is whatever the central entity in the situation seen through the telescope is. To comment on the view through the telescope is to evoke whatever entity is centrally imaged in the telescope as a prominent candidate for reference. And nothing else will do. Given the astronomical conjunction, the speaker couldn't distinguish Jupiter from Venus by pointing, nor could the visitor judge which body the telescope was pointed at by the direction of the tube. Letting s_1 name the view through the telescope, we can formalize the key bits of interpretation as follows:

(13) $\langle\alpha 0\rangle$; [Summary(s_0,e_0)]; $\langle\sigma 1 s_0\rangle$; $\langle\pi 1 @\textbf{that}\rangle$; [jupiter($e_0$, x_1)]
(14) $\langle\alpha 0\rangle$; [Summary(s_0,e_0)]; $\langle\sigma 1 s_0\rangle$; $\langle\pi 1 @\textbf{that}\rangle$; [venus($e_0$, x_1)]

The representations get the meanings right. More importantly, they explain how the visitor can recover the logical form and understand the speaker's point by recognizing the relationship that makes the speaker's utterance coherent, even though the visitor can't identify which specific body the speaker is referring to until the visitor looks through the telescope for herself. By contrast, if all you had were representations like (8), or some analogous representations of deixis that made reference explicit, you'd incorrectly predict that there's an ambiguity to resolve in (11) and (12) *even after* you understand them as comments about the view

through the scope. You'd have the two bright objects in the western sky, and you'd have to pick one and only one as the referent of the speaker's demonstration—or ask for clarification. But that's not how the interpretation proceeds. We take this as strong evidence against the idea that speakers and hearers must coordinate directly on demonstrative referents, a common view in both formal and computational semantics (Neale, 2004; Stone, 2004, King, forthcoming, a; forthcoming, b). What audience has to do is recognize the coherence relation that ties the utterance to the preceding discourse and the ongoing activity; but once this has been done, there's no need for further reasoning about speaker's referential intentions in order to resolve the reference of a demonstrative, beyond the fact that the speaker uttered *that sentence*. The reference resolution falls out of a linguistic mechanism associated with the coherence relation in question—for free.

Moreover, just as the audience can retrieve completely unambiguous information just by realizing that the speaker is making a comment about the particular situation, and relying on the world to disambiguate (e.g. by looking through the telescope), so too the speaker can contribute completely unambiguous information, even though she is not in a position to identify the object she is referring to directly herself. Let us explain. Suppose that the speaker walks into the room and sees a friend who has just deployed a telescope facing the western sky. The speaker asks 'What's that?' Clearly, before looking through the telescope, the speaker cannot directly identify which object she is referring to with 'that'; it could be any of the number of things on the western sky. Yet again, her question is unambiguous; all you have to do to understand her question is to figure out that she is making a comment about the scenery as seen through the telescope. Once you have done that there's no further ambiguity—the referent (if any) is the thing that's the central entity seen through the telescope. All the speaker and the hearer have to do is to look through the telescope and see for themselves which object that is. Or similarly, consider yet another telling example. Standing next to a massive sarcophagus, the local guide tells you "These are the relics of St. Rosalie, *La Santuzza*. The sarcophagus itself is made of pure silver." Of course the guide and the audience both unambiguously understand that the demonstrative 'these' refers to the content of the sarcophagus, while 'the sarcophagus' refers to the sarcophagus itself. Clearly, neither the guide nor the audience have actually opened the sarcophagus to witness what the referent of the demonstrative is. Yet they understand the contribution as unambiguous as soon as they recognize the speaker is commenting on the interior of the sarcophagus. Nothing less will do, and nothing more is needed. In principle, they could identify the referent just by looking inside.

5. Conclusion and Future Work

We have considered situated interpretations in coherent discourse, and argued that referential interpretations in cases like (1), (2) and (11) are understood

and derived relationally. This understanding requires representations of interpretations that explicitly associate discourse entities with the referents of pointing and other kinds of demonstration, and track the heterogeneous prominence that these entities get in virtue of the diverse relationships that utterances can bear to ongoing activity. In brief: we relate our talk of the world around us through suitable discourse relations.

Representing context dependence via mechanisms of discourse coherence provides an attractive framework in which to divide interpretation into stages and to minimize the problem solving that's necessary to compute logical form. Take (11) and its representation in (13). Here we have a formalism that captures the meaning of the utterance while spelling out the further work that will be required to resolve reference. According to (13), when you look through the telescope, you'll find out what the referent of 'that' is. Recognizing the logical form of the utterance this way should suffice for understanding. Moreover, we expect that the discourse relation could be resolved based on shallow constraints on what information counts as a summary. This reflects our broader expectation that disambiguation is generally a shallow process: while it does aim at recognizing the analysis of the utterance that the speaker had in mind, an aspect of the speaker's mental state, it is not the open-ended Gricean process of intention recognition that philosophers sometimes suppose.

Our approach commits us to representing utterances with specific kinds of interpretive connections to the world. Our characterization of these connections is obviously provisional, and corpus and modeling work is necessary to flesh out the parameters of the approach. It is worth pointing out that *Summary* and *Assessment* could also be used to formalize the interpretation of successive utterances by relating two described situations, not only relating an utterance to the real world situation. In fact, utterances can relate both to ongoing activity and to previous discourse. For example, consider (15) and (16), taken from Vi Hart's origami proof of the Pythagorean theorem—a visual narrative much like Child's where utterances describe ongoing events on the screen.[15]

(15) We're just taking advantage of the symmetries of the square for the next step.

(16) This is where you're choosing how long and pointy or short and fat the right triangle is.

Hart uses (15) while folding a square into eight identical segments to explain how to do the folds. Hart uses (16) as she describes the next step of folding, to highlight its result for the proof. Thus, these utterances are linked to the accompanying activity but do not just report what's going on; and they're linked to the ongoing discourse as well. In fact, Coherence Theory already allows that utterances can bear multiple connections to prior discourse (Asher & Lascarides, 2003). The closest parallel may be that of multimodal communication, described

in Lascarides and Stone (2009) who argue that utterances bear discourse relations both to prior utterances and to simultaneous gesture.

Despite the work that remains to be done, the main point remains—the relevant disambiguation in deictic utterances occurs not in disambiguating the meaning of the pronoun, but the type of coherence relation tying it to an accompanying real world situation. Once the relation is recognized, we argued, the interpretation of the pronoun falls out for free.

Notes

1. The example is a modification of an example from (Bolt, 1980).
2. Just which intentions are relevant for determining the referent of a demonstrative expression is a matter of considerable debate among the intetionalists. We look at different intentionalist accounts in more detail below.
3. It not controversial that number, person and gender constraints are linguistic constraints on reference of a demonstrative. Elsewhere (Stojnic, Stone & Lepore, 2013 ms.) we argue that demonstrative acts like pointings are likewise grammaticized, and thus provide linguistic constraints on reference.
4. For example, see King (2013; forthcoming a, b), Reimer (1992), Bach (1992), for different attempts on constraining naïve intentionalism.
5. It is instructive to consider an analogous case that does not involve demonstratives. Suppose the speaker intends to convey the information that Ann is happy, but a *lapsus linguae* occurs, and she ends up uttering "Sue is happy". No matter what she intended to convey, the content of her utterance is still that Sue is happy. The same is true in the case involving demonstratives.
6. Certainly, not the only way, though. The linguistic mechanisms that affect the structure of the context are diverse. For a more detailed account, see Stojnic, Stone & Lepore (2013 ms). Here we focus on discourse relations as the key mechanism in interpreting deictic utterances without overt demonstrative gestures.
7. See Kehler (2002) and Asher and Lascarides (2003) for more on discourse structure and its relationship to coherence relations.
8. For a more detailed development of these ideas, see Stojnic, Stone and Lepore, 2013 ms.
9. Conventional encoding of non-truth-conditional content is by no means unusual. For example, consider "Even I doubt myself sometimes", where 'even' contributes an implicature rather than truth-conditional content, but it still must be represented in logical form to be sensitive to the reflexive status of the predicate and the scope of 'sometimes'. For simplicity and for the purposes of this discussion, however, our formalism has only a single dimension of semantic content.
10. We use dynamic semantics because it provides an elegant framework for capturing our idea, but more importantly, because the full blown account of pronoun

resolution that takes into account all relevant anaphoric dependences will require encoding discourse dependences that are most straightforwardly captured in a dynamic framework. For more on the value of working in this framework, see (Stojnic, Stone and Lepore, 2013 ms).

11. Formally, we can think of a position in the sequence as analogous to a free variable that stores a value with respect to a discourse as a whole; less formally, they track contributions of noun phrases in a discourse (exactly the details of how this is done follow below). Note that even though we track the contribution of noun phrases *via* the mechanism of discourse reference, this does not compromise the key motivation behind the direct reference views, a la Kaplan (1989a), since discourse referents being free variables, preserve the intuitions about rigidity, that motivate the direct reference views. However, there is a sense in which discourse referents share the features of more fine grained, Fregean contents— namely, for any variables x and y, an agent can be ignorant of the value of x, but not y (and *vice versa*), even if *de facto* x and y have a same value. We welcome these features. Furthermore, appealing to discourse reference, in our view, offers the most plausible mechanism of capturing the discourse dependences relevant for accounting for donkey anaphora and cross-sentential anaphora, as well as more generally for the referential uses of expressions that normally do not have referential semantics, but nevertheless license anaphora. (Consider, for example: A: Put the cruise ship north of the Dominican Republic. B: It won't fit there.)

12. We use the term 'entity' in a neutral way. We shall see that we will need to recognize distinct type of entities (at least individuals, situations, events, and locations) in what follows.

13. To take a simple example, on this notation a representation of the sequence, "A man walked in. He sat down" would be:
i. $\langle \alpha 0 \rangle$; $\langle \alpha 1 \rangle$; [man(e_0, x_1)]; [walked.in(e_0,x_1)]; $\langle \alpha 0 \rangle$; [narration(@then^{e_0}, e_0,)]; $\langle \pi 1 @ he \rangle$; [sat.down($e_0$, x_1)], where '@then^{e_0}' denotes the top ranked event other than e_0.
The first line corresponds to the first sentence in our example. The first update introduces a discourse referent for an event, e_0, to the top position 0, and the update a discourse referent that is further constrained to be a man in e_0, to position 1. Moreover, it is required that the event is the event of walking in by the man. The second line corresponds to the second sentence in our example. The update introduces an event to position 0 (and as always, pushes all other referents one position down the stack) and requires that this event be related by Narration with the previous, just demoted one. Intuitively, the description of the second event continues the narrative that begins with the description of the first event (the event of walking in). Furthermore it is required that this event is the event of sitting down by the top ranked referent that is a third person, singular, male—that is, the man previously introduced. The interpretation is implicitly existentially generalized, so the discourse is true if and only if some man walked in and sat down, which is intuitively correct truth-condition.

14. For simplicity, we ignore the fact that the descriptions of the two entities are strictly speaking presupposed in this example. The treatment of presupposed information in dynamic semantics is beyond the scope of this paper.

15. http://www.youtube.com/watch?v = z6lL83wl31E

References

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Asher, N., & Lascarides, A. (2013). Strategic Conversation. *Semantics and Pragmatics* .

Austin, J. L. (1950). Truth. *Proceedings of the Aristotelian Society, Supplementary Volume*, 24, 111–128.

Bach, K. (2005). Context ex Machina. In S. v. Pragmatics, & Z. Gendler Szabo (Ed.): Oxford University Press.

Bach, K. (1992). Intentions and Demonstrations. *Analysis*, 52, 140–146.

Bach, K. (2001). You Don't Say. *Synthese*, 128, 15–44.

Barwise, J., & Etchemendy, J. (1987). *The Liar*. Oxford: Oxford University Press.

Barwise, J., & Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Blaylock, N. J. (2005). *Towards Tractable Agent-Based Dialogue (PhD thesis)*. Rochester.

Bolt, R. (1980). Put-That-There: Voice and Gesture at the Graphic Interface. *Computer Graphics*, 14(3), 262–270.

Clark, H. H., Krych, M., & A. (2004). Speaking While Monitoring Addressees for Understanding. *Journal of Memory and Language*, 50:62–81.

Crespo, I., & Fernandez, R. (2011). Expressing Taste in Dialogue. *SEMDIAL 2011: Proceedings of the 15th Workshop on the Seantics and Pragmatics of Dialogue*, (pp. 84–93).

Deker, P. (1994). Predicate Logic With Anaphora. In L. Santelmann, & H. M. (Ed.), *Proceedings From SALT IX* (pp. 79–95). Ithaca, NY: DMLL Publication: Cornell University.

Devlin, K. (1991). *Logic and Information*. Cambridge University Press.

Donnellan, K. S. (1966). Reference and Definite Descriptions. *The Philosophical Review*, 77, 281–304.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388.

Grosz, B. J., & Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3), 175–204.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2), 203–225.

Gundel, J. K., Hedberg, N., & R, Z. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69(2): 274–307.

Heim, I. (1982). The Semantics of Definite and Indefinite Noun Phrases, (Ph.D. thesis). Amherst: University of Massachusetts.

Hobbs, J. R. (1979). Coherence and Coreference. *Cognitive Science*, 3(1), 67–90.

Hobbs, J. R. (1985). Ontological promiscuity. *Proceedings of ACL*, (pp. 61–69).

Kamp, H. (1981). A theory of Truth and Semantic Representation. In J. A. Groenendijk, T. M. T.M.V. Janssen, & S. M. J. (Eds.), *Formal Methods in the Study of Language* (pp. 277–322). Amsterdam: Mathematical Centre Tracts 135.

Kaplan, D. (1989a). Demonstratives. In J. Almong, J. Perry, & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481–563). Oxford: Oxford University Press.

Kaplan, D. (1989b). Afterthoughts. In *Themes from Kaplan* (pp. 565–614.). Oxford: Oxford University Press.

Kehler, A. (2002). *Coherence, Refernce and the Theory of Grammar*. Stanford: CSLI.

Kehler, A., Kertz, L., H., R., & Elman, J. L. (2008). Coherence and Coreference Revisited. *Journal of Semantics*, 25(1), 1–44.

King, J. C. (2013). Supplementives, the Coordination Account, and Conflicting Intentions. *Philosophical Perspectives,* 27 (1).

King, J. C. (forthcoming a). Speaker Intentions in Context. *Noûs*.

King, J. C. (forthcoming b). The Metasemantics of Contextual Sensitivity. In T. N. Metasemantics, A. Burges, & B. Sherman (Eds.).

Kratzer, A. (1989). An Investigation of the Lumps of Thought. *Linguistics & Philosophy*, 12(5), 607–653.

Kratzer, A. (2002). Facts: Particulars or Information Units? *Linguistics & Philosophy*, 25(5–6), 655–670.

Kripke, S. (1977). Speaker's Reference and Semantic Reference. (P. A. French, T. J. Uehlig, & H. K. Wettstein, Eds.) *Midwest Studies in Philosophy, Volume II*, pp. 255–276.

Kruijff, G., Janicek, M., & H, Z. (2012). Situated Communication for Joint Activity in Human-robot Teams. *IEEE Intelligent Systems*, 27(2):27–35.

Lascarides, A., & Stone, M. (2009). Discourse Coherence and Gesture Interpretation. *Gesture*, 9(2): 147–180.

Lewis, D. K. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, 8(3), 339–359.

Lochbaum, K. E. (1998). A Collaborative Planning Model of Iintentional Structure. *Computational Linguistics*, 24(4), 525–572.

Luperfoy, S. (1992). The Representation of Multimodal User Interface Dialogues Using Discourse Pegs. *Proceedings of ACL*, (pp. 22–31).

Muskens, R. (1996). Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19(2), 143–186.

Neale, S. (2004). This, That, and the Other. In M. Reimer, & A. Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 68–182). New York: Oxford University Press.

Poesio, M., & Traum, D. R. (1997). Conversational Actions and Discourse Situations. *Computational Intelligence*, 13(3): 309–347.

Reimer, M. (1992). Three Views of Demonstrative Reference. *Synthese*, 93, 373–402.

Roberts, C. (2003). Uniqueness in Definite Noun Phrases. *Linguistics and Philosophy*, 26(3), 285–350.

Stojnic, U., Stone, M., & Lepore, E. (2013 ms.). Discourse and Logical Form.

Stone, M. (2004). Intention, Interpretation and the Computational Structure of Language. *Cognitive Science*, 28(5):781–809.

Voorhees, E. M. (2003). Evaluating Answers to Definition Questions. *Companion Volume of the Proceedings of HLT-NAACL – Short Papers*, (pp. 109–111).

Yu, C., & Ballard, D. H. (2004). On the Integration of Grounding Language and Learning Objects. *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, pp. 488–494.

Zeevat, H. (1999). Demonstratives in Discourse. *Journal of Semantics*, 16(4):279–313.