

5

Reviving the Parameter Revolution in Semantics

Bryan Pickel, Brian Rabern, and Josh Dever

... it is easy to write semantical rules that give the same analysis to recurrences of the same demonstrative (*what is hard is to write rules that don't*).

Kaplan (1989b: 590)

5.1 Introduction

Philosophy of language at the beginning of the twentieth century significantly advanced by examining formal languages used in mathematics. Working with these simplified languages, philosophers aspired to compositionally derive the meaning of a complex expression from the meanings of its parts and their mode of combination. But there were doubts about the prospects of extending this project to natural language due to its perceived imperfections such as the fact that the same natural language expression serves different linguistic functions in its different occurrences. Early semantic theorists had available only a limited range of tools to account for this variation. One prominent tool was positing ambiguity. Semanticists would translate natural language sentences into formal languages which resolve *lexical* and *structural* ambiguities. *Lexical ambiguity* arises when a single phonological string corresponds to distinct lexical entries or meanings. For instance, ‘bank’ is used to mean a financial institution or the shore lining a river. *Structural ambiguity* arises from unclarity in the mode of combination of the components of a compound expression.¹

Of course, ambiguity is a rather blunt instrument in accounting for an expression’s ability to serve different linguistic functions in different occurrences. A theory that posits additional lexical ambiguities thereby ascribes more knowledge to competent language users. This makes the requirements on learning the language more demanding. A theory that posits additional structural ambiguities attributes additional parsing powers to language users who understand the ambiguous sentences. This requires

¹ See Carnap (1937/1959: §46) on resolving lexical ambiguities in natural language by translating into a more precise symbolism. Russell (1905: 33) similarly discusses scope ambiguities.

them to have more computational abilities. For these reasons, it is desirable to avoid positing ambiguities unless necessary.² Consider an indexical pronoun such as ‘I’, which makes different truth conditional contributions in different contexts. A sentence containing this pronoun—such as ‘I am hungry’—may be true when uttered in one context, but false in another. If ‘I’ is ambiguous, then a distinct lexical entry is required for every speaker who uses it. If Ann and Bill are two individuals, then one must distinguish ‘I_{Ann}’ from ‘I_{Bill}’. Such pervasive ambiguity threatens to make language unlearnable.

Fortunately, Montague (1968), among others (e.g. Scott 1970; Lewis 1970; and Kaplan 1989a), began a revolution in semantics in order to account for the univocality of indexicals. Specifically, they treated context as a *parameter* at which a sentence is semantically processed. In a slogan, they taught us:

Parameterize, don’t lexicalize!

The truth conditions of a sentence in a context *c* are a function of the linguistic meanings of the components of the sentence evaluated at *c*. Thus, the sentence ‘I am hungry’ is true in a context *c* just in case the agent in *c* is hungry. This sentence is not ambiguous. Rather, it has a univocal but context-sensitive meaning which all competent speakers know.³ When Ann uses ‘I’, she refers to Ann. When Bill uses it, he refers to Bill.

But the revolution has stalled, and the threat of massive ambiguity reemerges. One salient problem comes from demonstratives. One may truly utter a sentence in a context *c* with two occurrences of a demonstrative such as (1), indicating a different individual with each occurrence of the demonstrative.

(1) He is tall and he is not tall.

But for (1) to be true at *c*, each occurrence of the demonstrative pronoun must make a different truth conditional contribution. The difference in truth conditional contributions cannot be accounted for by standard Kaplanian parameter sensitivity, according to which a sentence is assessed relative to a single context *c* and the truth conditional contribution of the demonstrative pronoun ‘he’ is a function of its linguistic meaning and *c*. For this reason, Kaplan (1989b: 586) himself posits ‘an

² Philosophers will be most acquainted with Grice’s (1975) principle known as modified Occam’s razor: ‘Senses are not to be multiplied beyond necessity.’ In syntax, Chomsky’s (1995) minimalist program is inspired partly by the desire to minimize processing requirements on language users. Computer scientists have assumed that natural language uses techniques to minimize computational complexity and have sought to incorporate these techniques into programming languages (see, e.g., Vermeulen 2000).

³ Kaplan (1989a: 505) says,

The character of an expression is set by linguistic conventions and, in turn, determines the content of the expression in every context. Because character is what is set by linguistic conventions, it is natural to think of it as *meaning* in the sense of what is known by the competent language user.

exotic kind of ambiguity, perhaps unique to demonstratives' whereby each occurrence of a demonstrative in a sentence has a different regimentation into a disambiguated language. On one of Kaplan's proposals, this disambiguation is effected by adding a numerical index to each occurrence of a demonstrative, seemingly making every occurrence of demonstrative its own lexical entry. Indeed, Gauker (2014) has recently argued that this puzzle concerning recurring demonstratives mandates a return to the massive ambiguity view: nearly every use of a demonstrative—and even of an indexical—is its own 'lexical item'.

Standard treatments of anaphoric pronouns, which are close relatives of demonstratives, wreak even more havoc, since they threaten to spread the ambiguity to other expressions of the language. Consider (2a) and (2b).

- (2) (a) An engineer saw an engineer.
(b) An engineer saw an engineer and she waved.

We assume that the sentence (2a) is univocal. However, the sentence (2b) has multiple readings, since, the pronoun 'she' may be interpreted as anaphoric so that its meaning depends on various linguistic antecedents. The linguistic antecedents in this sentence may be either occurrence of the determiner phrase 'an engineer'. The truth conditions of the sentence depend on which antecedent is selected. On one reading, the sentence is true just in case an engineer saw an engineer who waved. On another reading, the sentence is true just in case an engineer saw another engineer and waved.⁴

To handle this difference in interpretation, semanticists posit 'referential indices' on pronouns, following Chomsky (1965: 145). Referential indices are numbers assigned to occurrences of noun phrases in the logical form of a sentence. The different readings of (2b) emerge because the occurrences of 'her' are tagged with different referential indices. But, it's not enough merely to distinguish two meanings for the pronoun: 'she₁' and 'she₂'. A representation of the sentence which distinguishes these two meanings as in (2c) and (2d) does not distinguish the two meanings for the sentence as a whole.

- (2) (c) An engineer saw an engineer and she₁ waved.
(d) An engineer saw an engineer and she₂ waved.

To fully differentiate these two readings for (2b), semanticists also provide a numerical index for the determiner phrases which act as antecedents. Indeed, all noun phrases carry an index, according to this standard strategy. Thus, semanticists fully distinguish the two readings for (2b) as follows:

- (2) (e) An engineer₁ saw an engineer₂ and she₁ waved.
(f) An engineer₁ saw an engineer₂ and she₂ waved.

⁴ In order to emphasize the univocality of the antecedents, we have used cases of e-type pronouns. Similar cases can easily be constructed for pronouns that are c-commanded by their antecedents, as in: 'An engineer saw an engineer and her friend'.

Representations (2e) and (2f) generate the desired readings, but at a cost. Namely, in order to account for the diverse readings of (2b), which stem from the anaphoric pronoun, we have been forced to posit an ambiguity in the seemingly univocal ‘an engineer’ and thus also in the seemingly univocal (2a), which is resolved by decorating each determiner phrase with an index. The contagion has spread from anaphoric pronouns to all noun phrases in natural language. These are now treated as massively ambiguous.⁵

To bring the parameterization revolution closer to completion, we propose a semantics that overcomes these entrenched pockets of resistance. In particular, our semantics delivers the following two results.

DEMONSTRATIVE UNIVOCALITY: A demonstrative need not be ambiguous in order to account for the different truth conditional contributions of its different occurrences.

ANTECEDENT UNIVOCALITY: The antecedent for an anaphoric pronoun need not be ambiguous in order to account for different possible anaphora resolutions.

DEMONSTRATIVE UNIVOCALITY requires that the two occurrences of pronouns ‘he’ in (1) ‘he is tall and he is not tall’ are assessed at different contexts. We propose to

⁵ The problem generalizes what Fine (2003, 2007) calls *the antinomy of the variable* (cf. Jacobson 1999: 127). In particular, quantifiers seem to be univocal. But they also seemingly make different truth conditional contributions by binding different variables. There is a temptation to think that an existential quantifier such as ‘an engineer’ is univocal in all of its occurrences. Thus, philosophers in the Fregean tradition might be tempted to treat this expression as designating a second-level property possessed by first-level properties which are instantiated by engineers. On this view, ‘ $\exists x$ ’ and ‘ $\exists y$ ’ would symbolize the same second-level property. But even Frege himself needed to syntactically differentiate the various occurrences of a quantifier. In modern symbolism, this result is achieved by the fact that the quantifier ‘ \forall ’ is connected to distinct variables ‘ x ’ and ‘ y ’ with these variables making distinct truth conditional contributions. (See Rule 2 of Frege (Frege et al. 1893/2013: §8). For example, given that the function $\xi = \xi$ has the True as value for every argument we form the corresponding expression of generality as ‘ $\underbrace{\xi}_{\alpha} = \alpha$ ’, but given that the function $\underbrace{\xi}_{\alpha} < \alpha$ has the True as value for every argument, it’d clearly be a mistake to conclude that $\forall x \exists x x < x$, which we might try to express in broken *Begriffsschrift* as ‘ $\underbrace{\underbrace{\xi}_{\alpha}}_{\alpha} < \alpha$ ’. Rule 2 (§8) mandates that the corresponding expression of generality must choose a distinct German letter ‘ $\underbrace{\underbrace{\xi}_{\alpha}}_{\epsilon} < \alpha$ ’.) Barwise and Cooper (1981) implement quantified NPs as second-order properties (sets of sets, strictly) in a such a way that at first glance makes it look like they get away without indices. The roles of quantification and variable binding are separated, and the explicit story about variable binding is suppressed. Their syntax (§2.3) includes the abstraction operator that combines with a formula to form a set term (rule R2), but they then omit the semantics for the abstraction operator—it would clearly have to include explicit clauses concerning variable co-indexing (cf. Lewis 1970: 45, and Heim and Kratzer 1998: 186). So while it is true that the quantifiers in their formal language such as ‘some(thing)’ are univocal, the translation of a natural language sentence containing a quantified noun phrase, will include an (indexed) abstraction operator that merges with the formula embedded under the quantifier. For example, consider their representation of ‘Most men kiss (particular) women’: **some(women)** \hat{y} [**most(men)** \hat{x} [**kiss** (x, y)]]. Analogously, Heim and Kratzer (1998: §10.2) posit univocality for determiner phrases (DPs) other than pronouns so that ‘[f]or non-pronominal DPs, indices are in principle optional’. But this results from the fact that sentences containing these DPs are construed as also containing covert λ -binders which are co-indexed with any pronouns they may bind. Moreover, Heim and Kratzer go on to offer a standard argument for indexing non-pronominal DPs from movement, which presupposes that moved constituents leave empty traces in their moved positions.

evaluate expressions against both an extra-linguistic context and a discourse context in order to determine their extensions. On our account, the two occurrences of ‘he’ are evaluated at the same extra-linguistic context, but different discourse contexts. In the processing of a sentence, the discourse context evolves by tracking the occurrences of demonstratives and other noun phrases as they are used. The n^{th} demonstrative refers to the n^{th} demonstrated object in the context. So sentence (1) is true in a context c just in case the first demonstrated object in c is tall and the second demonstrated object in c is not tall. As a result, demonstratives are context-shifting expressions, on our account. Turning to ANTECEDENT UNIVOCALITY, we generalize this framework so that the discourse context tracks occurrences not only of demonstratives, but of all potential anaphoric antecedents in a uniform way. Discourse context then provides a rich enough structure to link anaphoric pronouns to their antecedents without rendering these lexically ambiguous—the background syntactic representations are a version of the devices employed by De Bruijn (1972) for binding relations in the lambda calculus.

5.2 Ambiguity Strategies

A single demonstrative may recur within a sentence, making distinct truth conditional contributions. Thus, if one points to different objects as one pronounces each occurrence of ‘that’ an utterance of (3) may be false.

- (3) That is identical to that.

For the sentence to be false, the two occurrences of ‘that’ must refer to different objects and so make different truth conditional contributions. This difference in truth conditional contributions cannot be accounted for by simple parameter sensitivity, according to which a sentence is assessed relative to a single context c and the truth conditional contribution of the demonstrative pronoun ‘that’ is a function of its linguistic meaning and c .⁶

Spelling this out explicitly (D1)–(D4) are inconsistent.

- (D1) The sentence ‘that is identical to that’ is false in some contexts.
- (D2) The truth-conditional contribution of an occurrence of the word type ‘that’ used as a demonstrative is a function of the occurrence’s linguistic meaning and the context of use.
- (D3) The truth-conditions of a sentence in a context c are determined by the truth-conditional contributions of the occurrences of its constituent expressions relative to c and their mode of combination.

⁶ This problem goes back to the early versions of Kaplan (1989a), and is discussed by Lewis (1970: 62):

[C]onsider the sentence “*This* is older than *this*”. I might say it pointing at a 1962 Volkswagen when I say the first “*this*” and at a 1963 Volkswagen when I say the second “*this*”. The sentence should be true on such an occasion; but how can it be?

- (D4) Every occurrence of the word type ‘that’ (as a pronoun) has the same linguistic meaning, and thus every occurrence of the sentence type ‘that is identical to that’ has the same linguistic meaning.

An account of recurring demonstratives must reject one of these claims. Many accounts on offer reject (D4), the univocality of demonstratives, in some form. Our complaint is that, while the slogan—*Parameterize, don’t lexicalize!*—has been taken on board in *some* regions of theorizing about context sensitivity (‘I’, ‘now’), it has been neglected in favor of continued ambiguity theorizing in others—especially in the semantics of demonstratives and pronouns.

In this section we do two things. First, we make the case that people have in fact been lexicalizing, not parameterizing. This isn’t immediately obvious. There are numerous semantic proposals for recurring demonstratives and for anaphoric pronouns. Many of these proposals make use of some Kaplan-style machinery that extracts semantic values relative to a contextual parameter. (One of the proposals is even Kaplan’s own proposal.) It would thus be easy to take these proposals as parameterization proposals. But they are not, at least not thoroughly—we will bring out the central role that lexicalization of ambiguity plays in these various proposals.

Once we’ve made the case that standard packages for handling recurring demonstratives are best seen as ambiguity approaches, readers might react by suspecting that ambiguity approaches are not so bad after all, and that thorough-going parameterization was just a pipe dream. The second thread of our backing up slogan-plus-complaint is thus to remind readers of the costs of ambiguity strategies, and to trace out the manifestations of those costs in the standard approaches to recurring demonstratives. What we say here isn’t intended as a *refutation* of ambiguity approaches. We are skeptical that ambiguity approaches can be refuted in any very strong sense. It is surely possible to set out an ambiguity treatment of *prima facie* context-sensitive language; the most that can be said is that such ambiguity treatments will typically be *ad hoc* and underexplanatory, and hence that an ambiguity-free approach, should one be possible, would be preferred. Later in the chapter we take on the burden of providing a possibility proof.

Strategy 1: Lexical multiplication. We begin by considering, as an initial stalking horse, the story mentioned above on which English comes with a large collection of type-distinct first-person pronouns, so that for each individual there is a first-person pronoun type that refers (in a context-insensitive way) to that individual. (One might attribute this view to Rodgers and Hammerstein with their lyric: ‘*me*’, *a name, I call myself*.) This story captures the referential variability of the word ‘I’ through lexicalized ambiguity, rather than by the familiar Kaplanian strategy of assigning reference only relative to a (contextual) parameter. Why might we prefer a Kaplanian parameterized treatment of ‘I’? Some considerations:

1. Because there are many speakers, an ambiguity account will require many first-personal pronouns (one for each speaker). This proliferation of the lexicon is troubling twice over:
 - (a) *Learnability Objection*: It makes speakers' ability to learn the language mysterious, both because it vastly increases the numerical lexical burden and because it requires speakers regularly to know the meaning of words they have never encountered before.⁷
 - (b) *Explanation Objection*: It fails to capture an important linguistic generalization connecting the producer of a first-person pronoun with the referent of that pronoun, and thereby diminishes our ability to give powerful linguistic explanations.
2. *Communication Objection*: The ambiguity theory creates spurious communicative options. Because the ambiguity theory posits one word 'I' referring to Carnap and another word 'I' referring to Church, it predicts that *Carnap* can say 'I am a logician' in order to claim that *Church* is a logician. But this is not a genuine communicative option for Carnap.⁸

A similar story can be told about demonstratives.⁹ We could take English to be stocked with a large (very large—one for every possible demonstratum) collection of demonstrative types, each of which refers (in a context-insensitive way) to a specific object. Where $\{A, B, C, \dots\}$ is the class of possible demonstrata, we would then have a long list of lexical entries as follows:

$$\begin{aligned} \llbracket \text{that}_A \rrbracket^{c,w} &= A \\ \llbracket \text{that}_B \rrbracket^{c,w} &= B \\ &\dots \end{aligned}$$

To further extend the story to anaphoric pronouns, we would need English to be stocked with a collection of pronoun types, with one type for each potential antecedent noun phrase.

This ambiguity story about demonstratives is subject to the same worries as the ambiguity story about indexicals. If anything, the concerns are heightened in this case, due to the increased cardinality of the ambiguity. Again we are left with a picture of

⁷ Perhaps the learnability burden is not so high as it might initially appear, because language users can plausibly inductively determine the pattern that each speaker uses a first-person pronoun that refers to them?. But this looks like a re-introduction of the parameterized Kaplanian semantic value. (Although matters are subtle here. Does an expectation that the fifth child will be named Quentin amount to linguistic mastery of a parameterized semantic value, or just to a culturally-informed guess?)



⁸ Of course, the ambiguity theory can be supplemented with additional epicycles to block the spurious options. There can, for example, be a pragmatic principle that no speaker uses any first-person pronoun other than the one that refers to themselves. But the predictive and explanatory power of the theory will again be reduced by the reliance on such epicycles.


⁹ We don't claim that anyone actually has told this story, although as noted above the view endorsed by Gauker (2014) is along these lines.


language that makes acquisition mysterious, that fails to capture generalizations about the many words spelled ‘t-h-a-t’ and ‘t-h-i-s’, and that allows speakers to say things that don’t, in fact, seem say-able (by, e.g., using demonstratives that refer to objects that aren’t in any sense salient in context).





Strategy 2: Completion by demonstration. On one way of taking Kaplan’s view on demonstratives, that view is something close to the ambiguity theory just sketched. (We will present a second way of taking Kaplan’s view shortly.) Kaplan tells us that ‘a demonstration accompanies every demonstrative and determines its referent’ (p. 585). One way that a demonstration can *accompany* a demonstrative is by actually being *part of* the demonstrative. So on this way of reading Kaplan, there are in any important sense *many* demonstratives, one for each demonstration. There is, perhaps, only a single ‘lexical’ item ‘t-h-a-t’, but that item is no longer by itself a demonstrative. Rather, it is a component of an expression combining ‘t-h-a-t’ with a demonstration to form a full demonstrative—strictly ‘t-h-a-t’ only occurs as an ‘orthographic accident’. There will be many of these ‘that’-demonstration pairs. If demonstrations are then individuated by their demonstrata, we have one demonstrative for each potentially demonstrated object. Consider the sentence (4).

(4) He walks.

(4) may be uttered while demonstrating Carnap or while demonstrating Church. Let these demonstrations be ‘’ and ‘’, respectively. Then we have the following representations to be evaluated for truth.

(4.1) He-] walks.

(4.2) He-] walks.

The difference in truth conditional contribution between the two demonstrative uses of ‘he’ derives from a lexical ambiguity. It is not really ‘he’ alone that we are using, but rather ‘He-’ and ‘He-’. When a demonstration is of Carnap, the bundled representation ‘He-’ refers to Carnap, but when a (different) demonstration is of Church, the different bundled representation ‘He-’ refers to Church. This would thus require a long list of lexical entries for bundled ‘that’-demonstration pairs as follows:

[[that-]]^{c,w} = A

[[that-]]^{c,w} = B

...


The resulting proliferation of demonstratives leads Salmon to say:

On Kaplan’s theory . . . each utterance of ‘that’ with a different designatum is an utterance of a different term with a different character or meaning. . . . One might say that the demonstrative ‘that’ is highly ambiguous on Kaplan’s account, its precise meaning depending on the content of

the accompanying demonstration. This is not merely somewhat counterintuitive; it is obviously incorrect. (Salmon 2002: 512)

This version of Kaplan's theory shares the troubling features of the massive ambiguity story about 'I'. Language learners are left with a vast (probably infinite) primitive vocabulary to learn. And because there are separate reference axioms for each demonstrative (or 'that'-plus-demonstration pair), no explanation is provided for the semantic commonality among demonstratives.

Finally, too many communicative options are left open for speakers (and thus too many interpretive options are left open for audiences). On this view, any speaker is free to produce an utterance of 'That is tall' with 'that' referring to the Eiffel Tower, or to the Empire State Building, or to Mount Everest.¹⁰

One could insist that each 'that'-plus-demonstration pair is actually a complex expression: 'that' and the demonstration are each lexical items which compose to form the complex referring expression 'that '. In this way, 'that' itself can be supplied with a univocal meaning (the identity function), while each demonstration requires a separate lexical entry.

$$\llbracket \text{that } \langle \text{pointing hand} \rangle \rrbracket^{c,w} = \llbracket \langle \text{pointing hand} \rangle \rrbracket^{c,w} = A$$

$$\llbracket \text{that } \langle \text{pointing hand} \rangle \rrbracket^{c,w} = \llbracket \langle \text{pointing hand} \rangle \rrbracket^{c,w} = B$$

...

This preserves the univocality of 'that' only by putting all the action on the demonstrations—the demonstrative alone does not suffice for determination of a referent (even relative to a context). Thus, this just postpones the issue. Language learners are still left with a vast primitive vocabulary to learn. And in this case, the vocabulary consists largely of demonstrations—pointing fingers, gestures, glances, directing intentions, must all be construed as lexical inputs to interpretation. (Consider the analogous story in the case of 'I', whereby the input to interpretation is a complex made up of an identity function and a speaker. Since this doesn't avoid the proliferation of the lexicon, nor account for the lack of certain communicative options, it doesn't ultimately avoid the objections.)

On a modified version of this view demonstrations are not individuated by their demonstratum, but rather by the descriptive content of the demonstration. This is thus equivalent to stocking the language with an infinite collection of demonstratives, each of which has the logical form of ' $\text{dthat} - [\text{the } \pi]$ ', for some predicate π . (This is plausibly Kaplan's actual view in Kaplan 1989.) Since on this view a demonstrative

¹⁰ Perhaps this is too quick, because speakers in many contexts won't be able to perform demonstrations of all of these objects?. Two responses. First, it would be unfortunate if an overly generous picture of speaker communicative options was reined in only because some utterances were, as it were, too hard to pronounce. Second, to make this move is in effect to switch to a parameterized version of Kaplan's view, on which demonstrations have content relative to a context. (And then to say that in some contexts, no demonstration demonstrates the Eiffel Tower.) We consider the parameterized version of Kaplan below.

has a referent only relative to a parameter—relative to a context c , a demonstrative refers to whatever object satisfies the descriptive content of the demonstration in the world w^c of the context. Thus the same demonstrative can refer to different objects in different contexts.

$$\llbracket \text{dthat} - [\text{the } F] \rrbracket^{c,w} = \text{the } F \text{ in } w^c \text{ (if there is a unique one)}$$

Parameterization offers an alternative to lexicalized ambiguity, so one might think that this version of Kaplan's view avoids the problems above. Yet, this view remains too timidly parametric. In order to get non-coreferential demonstratives, in a single context, we need two different demonstrative types, achieved by bundling 'that' with two different demonstrations—that is, two demonstrations with different descriptive content. We distinguish 'dthat-[the F]' from 'dthat-[the G]', and 'that' paired with infinitely many other 'descriptive pseudodemonstrations' as well.

$$\llbracket \text{dthat} - [\text{the } F] \rrbracket^{c,w} = \text{the } F \text{ in } w^c$$

$$\llbracket \text{dthat} - [\text{the } G] \rrbracket^{c,w} = \text{the } G \text{ in } w^c$$

...

So even with parameterization added, an element of ambiguity is still required to handle the problem of recurring demonstratives. The continued element of lexical ambiguity means that we continue to confront the serious disadvantages of lexicalization strategies. The learning burden on language users remains high, since speakers still need to learn an infinite number of demonstratives. Note that it's important here that the descriptive content of demonstrations is *covert*. If a language—such as Kaplan's formal language LD—actually used phrases of the form $\lceil \text{dthat } \alpha \rceil$, then speakers could exploit their finitely grounded compositional mastery of the semantics for an infinite number of α to have a comprehensible learning strategy for the full range of demonstratives. But when language learners are only ever confronted with the single morphological type 'that', there is no such convenient story about learning.

A second disadvantage of both versions of the Kaplan strategy is that neither provides resources for handling anaphoric uses of demonstratives. Treating true demonstratives as a combination of the word 'that' with a demonstration handles demonstratives picking out objects deictically, but not demonstratives anaphorically linked to other noun phrases in discourse. Kaplan's approach is simply to separate the two cases and treat anaphora as a separate phenomenon. This is obviously less satisfactory than a unified treatment, and in any case the needed treatment of anaphora threatens further invocation of ambiguities.

Strategy 3: Numerical indexing. A more sophisticated appeal to parameterization does permit a unified treatment of demonstrative and anaphoric pronouns. What is needed is a device which can connect a demonstrative either (a) to a parametrically provided object or (b) to a previous discourse element on which the demonstrative is anaphoric. The standard method of introducing such devices is to 'tag' both

demonstrative and antecedent noun phrase occurrences with *indices* (see, e.g. Fiengo and May 1994 and Heim and Kratzer 1998). The indices then play the role of associating the tagged demonstrative either (a) with an appropriate component of the parameter¹¹ or (b) with an antecedent noun phrase tagged with the same index, which will then enter into a content-providing relation to the demonstrative.

Because there are many objects available for demonstration and many noun phrases available as antecedents, an indexing approach needs many indices. Demonstrations, at this level of abstraction, could be treated as indices—the difficulty here, which makes Kaplan’s demonstration-based semantics ill-suited for dealing with anaphora, is that there is no plausible account of the way in which antecedent noun phrases are tagged with indices of this sort.¹² Thus Kaplan’s account *can* be viewed as an indexing story. However, indexing strategies are most commonly implemented by allowing indices to be numerical rather than arbitrary objects.¹³

Numerical indexing conveniently provides an unlimited source of indices (needed, because there is no upper limit on the number of anaphoric or demonstrative pronouns a sentence can contain), together with a simple mechanism by which an indexed expression is associated with a value at a parameter. Following Tarski (1936), one may think of the parameters as sequences. On this view, the truth conditional contribution of an expression bearing the i^{th} index is determined by how it affects the i^{th} member of the evaluation sequence. Thus, two expressions bearing the same index will make the same truth conditional contribution at the same parameter, while two expressions bearing different indices may make different truth conditional contributions at the same parameter (cf. the ‘linking rule’ of Fiengo and May 1994). Letting σ^c be the sequence of demonstrata supplied in a context c we have the following lexical entries:

$$\llbracket \text{that}_1 \rrbracket^{c,w} = \sigma_1^c$$

$$\llbracket \text{that}_2 \rrbracket^{c,w} = \sigma_2^c$$

...

Indexing thereby purportedly solves the problem of recurring demonstratives, since multiple occurrences of the pronoun ‘he’ in (1) ‘he is tall and he is not tall’ may bear

¹¹ Our parameters will thus need to link objects with indices. As we see below, a standard way to do this is to take indices to be numbers and parameters to be (implicitly numbered) sequences of objects.

¹² It is, we hope, clear enough that in an utterance of ‘A linguist gave her talk’ in which ‘her’ is bound by ‘a linguist’, there is nothing in surface form that amounts to a *demonstration* linked to ‘a linguist’, so the indexing here is objectionably covert. One of our points then is that the indexing is often equally objectionably covert with deictic demonstratives—in saying ‘That is my point exactly’, there need be nothing in surface form that amounts to a demonstration, especially in the *dthat* sense—and that the indexing of antecedent noun phrases is no less objectionably covert when we move to numerical indexing.

¹³ The use of numerical indices is standard in most formal semantic traditions—see, for example, Heim and Kratzer (1998). Kaplan (1989a) also considers attaching subscripts to demonstratives and letting the n -th demonstrative designate the n -th demonstratum in a context, but he preferred the analysis in terms of ‘*dthat*’ for its alleged epistemological virtues (cf. 528–9).

distinct indices. If the indices are numerical, then one may regiment this sentence as (1*).

(1*) He_1 is tall and he_2 is not tall.

Relative to a contextually supplied sequence, ' he_1 ' will contribute the first member and ' he_2 ' will contribute the second. Thus, the two occurrences of 'he' may contribute different individuals because they correspond to distinct underlying structures.

As above, we distinguish two readings of 'An engineer saw an engineer and she waved' by differentially distributing indices:

- (2) (e) An engineer₁ saw an engineer₂ and she₁ waved.
 (f) An engineer₁ saw an engineer₂ and she₂ waved.

Note again that it is crucial that indices accompany antecedent as well as anaphor, to get both ends of the coordination accomplished.¹⁴ The indexing of the antecedent will not have any local role in the semantic interpretation of the antecedent, so there are extra difficulties in getting a theory of antecedent indexing properly constrained by the data.

Indexing strategies, of course, require indices. Infinitely many indices, if we pursue a numerical indexing strategy. And the use of infinitely many indices continues to raise the Learnability, Explanation, and Communication concerns about ambiguity strategies.¹⁵ Some of these concerns are partially mitigated by the use of numerical indexing. The ordinal structure of the numerical indices provides a systematic connection to parametric positions, and thus provides a somewhat more explanatory account. Even though there are infinitely many indices, there is a plausibly learnable systematic semantic story on which the n^{th} index serves to pick out the n^{th} element of the parametrically-provided sequence.¹⁶ Still, the Communicative difficulties remain. Numerical indices allow for the production of utterances such as 'That₂₇ is that₁₃₁', which states that the 27th most salient object in context is identical to the 131st most

¹⁴ All that matters here is relations of sameness and difference among the indices, not their ordinal positions. That is, the specifically *ordinal* aspect of indices is not exploited in this treatment of anaphora. This fact points the way to linking approaches, discussed below. Our own positive account, on the other hand, will exploit ordinal features in treating anaphora, but will *not* require ordinal features in the treatment of demonstratives.

¹⁵ Advocates of the Minimalist Program in syntax also standardly reject indices for similar reasons. In particular, they adopt a very strong syntactic principle, the *inclusiveness constraint*: (Chomsky 1995: 228). 'No new information can be introduced in the course of the syntactic computation' (Radford 2004: 94). (See (e.g.) Safir 2004 and Rooryck and vanden Wyngaerd 2011.) Our argument, however, requires nothing so strong as the inclusiveness constraint.

¹⁶ One way to think about this is to follow Lewis (1970: 62ff) in generating all of the necessary indices by successive applications of a single operation corresponding to successor. Thus, the first variable might be ' x ', the second variable ' x'' ', and so on. On this approach, the base producing the indices is in fact finite, and allows a recursive semantic theory. (To proceed in this way is in essence to trade off lexical ambiguity for structural ambiguity, since ' $that_1$ ' and ' $that_2$ ' now have different syntactic structures, with the tree of the latter branching one level deeper than the tree of the former.)

salient object in context. But this doesn't seem to be a genuine communicative option, and if it were, it is hard to see how language learners could become aware of the availability of such options.

Indexing strategies can be thought of as preserving DEMONSTRATIVE UNIVOCALITY, but only in the thinnest of senses. The indexer can claim that 'he' is unambiguous, because the ambiguity is carried by the index '1'. But to say this is also to say that the demonstrative alone does not suffice for determination of a referent (even relative to a context), and this price is at least as high as the price of denying DEMONSTRATIVE UNIVOCALITY. Better to say that the demonstrative proper is the concatenation of the overt demonstrative and the index, in which case referential determination is preserved and DEMONSTRATIVE UNIVOCALITY is lost. And for indexing strategies, as goes DEMONSTRATIVE UNIVOCALITY, so also goes ANTECEDENT UNIVOCALITY. Since semantic coordination is achieved by sameness of index, the only tool for getting anaphoric pronouns coordinated with their binding antecedents is to decorate both with the same index, which of course requires decorating the antecedent. Multiple decorations are available, so the antecedent (proper) becomes ambiguous.

Strategy 4: Syntactic linking. Numerical indexing strategies fall prey to our concerns about lexicalizing ambiguity approaches because they posit covert indices which carry a crucial part of the semantic burden. A final strategy attempts to offload the burden from the lexicon to the syntax, by making use of *structural* rather than *lexical* ambiguities.¹⁷ It is easiest to articulate such an approach in the case of anaphora. Consider a standard case of anaphora resolution such as (5).

- (5) When a man sees his father, he asks him for money.

Sentence (5) may be taken as asserting that when a man sees his father, the man asks the father for money. Alternatively, it may be taken as asserting that the father asks the man for money. The ambiguity is resolved by determining the anaphoric relations between the anaphoric pronouns, 'he' and 'him', on the one hand and their antecedents, 'a man' or 'his father', on the other.

On the views under consideration, these ambiguities are structural because they are to be resolved by determining the *syntactic relations* between the anaphoric pronouns and their antecedents. That is, the anaphoric pronouns 'he' and 'him' have the same meaning in (5). The ambiguity arises because of the relations these expressions stand in to their possible antecedents. Following Evans (1977) and Higginbotham (1980, 1983), one might represent the two inputs to semantic processing using arrows as follows.¹⁸

¹⁷ One possible additional motivation for this shift: indexing strategies by their nature create *symmetric* relations between co-indexed items. But as Higginbotham (1980, 1983) observes, anaphoric pronouns *asymmetrically* depend on their antecedents. The kinds of syntactic 'linking' relations we now consider can easily be taken to be asymmetric.

¹⁸ Quine (1940/1981: 69–70), suggested similar devices with relations of variable binding represented using 'quantificational diagrams', where lines or 'bonds' connect quantifiers to the positions in predicates

(5.1) When a man sees his father, he asks him for money

(5.2) When a man sees his father, he asks him for money

In (5.1), ‘he’ is anaphoric on ‘a man’ and ‘him’ is anaphoric on ‘his father’. In (5.2), ‘he’ is anaphoric on ‘his father’ and ‘him’ is anaphoric on ‘a man’.

This approach could also be extended to cover the case of recurring demonstratives. Grammatical relations between expressions and elements of extralinguistic reality could give rise to structural ambiguities of the same sort as we saw with (5). Consider sentence (4). Some have posited that (4) is *structurally ambiguous*, giving rise to two different representations that are evaluated for truth. The different representations are made explicit by exhibiting quasi-anaphoric dependence between the pronoun ‘he’ and different accompanying demonstrations. Thus, (4) may be uttered while demonstrating Carnap (‘☞’) or while demonstrating Church (‘☞’). Then we have the following representations to be evaluated for truth.

(4.3) ☞ He walks

(4.4) ☞ He walks

Thus, the difference in truth conditional contribution between the two demonstrative uses of ‘he’ arises from the same source as the difference in truth conditional contribution between the two anaphoric pronouns in (5). In a sentence containing multiple occurrences of a demonstrative, the demonstratives might bear linking relations to different elements of extralinguistic reality and thereby make different truth conditional contributions.¹⁹

Linking strategies can give the appearance of preserving DEMONSTRATIVE and ANTECEDENT UNIVOCALITY by removing the coordinating devices from the lexical items themselves and relocating them into the higher-level syntactic structures. But the appearance is deceptive. It is not enough for our linguistic resources to contain a single lexical type ‘a man’ and a single lexical type ‘he’ together with a linking ‘wire’ intended to reveal the binding relation between the two. We also need to get these

that they bind. This idea is also echoed in Kaplan (1986: 244), who alludes to the connection to Frege’s syntax whereby ‘variables’ (i.e. German letters) are merely typographic parts of the quantifier sign serving to link the concavity to the relevant ‘gaps’ in predicates.

¹⁹ A similar use of incorporation of *prima facie* extralinguistic reality into utterances can be found in Fine (2007: 124), who explicitly endorses a treatment of deictic pronouns as exhibiting a sort of ‘anaphoric’ dependence on extralinguistic reality, namely as ‘anaphoric on an associated demonstration’. This approach might make use of the sort of ambiguity considered here. Hunter (2013; 2014), and Stojnic et al. (2013) also model deictic pronouns on anaphoric pronouns. However, their discussions are embedded within the context of *Discourse Representation Theory* (DRT).

pieces assembled in the right way. So we need ‘a man’ to be associated with (one end of) the linking wire and ‘he’ to be associated with the other. This is just a re-emergence of the starting problem. We thus need a distinction between ‘*him*’ *linked to a wire linking to ‘a man’* and ‘*him*’ *linked to a wire linking to ‘his father’*. One way to think about this is that we return here to the indexing strategy, and take the semantically relevant units to be bundles of lexical items and wires, rather than lexical items alone and wires alone. Alternatively, the lexical indexing strategy can be thought of as a version of the structural linking strategy, with the indices serving as a lexical collection large enough to run a simple parameterization strategy on and then those indices wired to the overt demonstratives via structural linkings. On examination, the space between the indexing and the linking approaches diminishes, but both are unambiguously forms of ambiguity strategies.²⁰

Summarizing. We have examined a number of approaches to treating the differing patterns of reference and coreference with demonstratives and anaphoric pronouns. These approaches all share an aspiration to explain the semantic features through some degree of lexicalization (or, in the case of linking, grammaticization). Some of these strategies offload *part* of the explanatory burden to a partial role for parameterization, but the parameterization is always accompanied with massive ambiguity. Such approaches thus incur significant syntactic commitments, commitments that run counter to current thinking in syntax. And no such approach deals adequately with the Learnability, Explanation, and Communication objections. The goal moving forward is thus to find the *right* way to parameterize.

5.3 Towards Context Shifting

We aim for a theory that preserves (D4), a sentence such as ‘that is identical to that’ is univocal in the sense that it has a single context-dependent meaning. In particular, any two uses of the demonstrative have the same linguistic meaning and there is no ambiguity in the sentence’s structure. Since some utterances of ‘that is identical to that’ are false, the two occurrences of ‘that’ must make different truth conditional

²⁰ The earlier concerns of footnote 15 about the compatibility of syntactic theory with indexing strategies reoccur with linking strategies. If links are syntactic relations other than concatenation (e.g. in the Minimalist program, other than whatever relation is created by application of Merge), then the Inclusiveness Constraint again creates difficulties. This constraint seemingly contradicts any syntactic implementation of the linking strategy, which is why many of its proponents now suggest that the linkage only arises in the semantics or propose to find ways to weaken the inclusiveness constraint. For example, Safir (2004: 44ff) makes the following remark:

[O]ne might conclude that dependency relations are not in syntactic representations, but only in semantic representations. As relations that are functionally necessary for understanding, their existence is a bare output condition, which any computational solution to the relation of form and meaning must permit. The use of arrows, then, is merely descriptive notation, but not part of syntactic representation itself.

contributions. But since these occurrences have the same linguistic meaning, these occurrences must be evaluated at different contexts. This would require some kind of mid-sentence *context shift*—in the semantic assessment of a sentence different parts of the sentence must appeal to different contexts.

We will first argue (§5.3.1) that the context shift must be controlled, rather than uncontrolled, since semantics should provide truth conditions for each sentence at a context. We then explore (§5.3.2) one of the best extant theories of how the context evolves when there are multiple demonstratives, defended by Braun (1996). In this theory, a demonstrative shifts the extralinguistic context by making salient the next demonstratum (or demonstration). We argue that Braun's theory errs by trying to handle discourse effects solely using extralinguistic context. On our view, the shift induced by a demonstrative should be modeled as a change in *discourse context*. We will show the problems in Braun's theory that result from assuming otherwise.

5.3.1 *Uncontrolled context shifting*

In a false utterance of the sentence 'that is identical to that' distinct objects will be demonstrated with each occurrence of 'that'. For example, one may demonstrate a book with the first occurrence and a hat with the second occurrence of 'that'. It's natural to think that the difference in demonstrations entails that the expressions should be evaluated with respect to different contexts c and c^* . In c , a book is the salient demonstratum. In c^* , a hat is the salient demonstratum. This idea can be motivated by considering the fact that during a conversation the 'context' can *change*. The time moves on, the person speaking changes, the salient objects change, etc. A goat might walk into the room, thereby making a new object salient. In these ways the context seems to 'move beneath our feet'. We can exploit these changes to talk in a seemingly truthful manner, e.g. when the magician says 'Now you see it but *now* you don't' or when spouses finish each other's sentences 'I am tidy *but I'm not*'.

Is there a way to provide a compositional semantics that accommodates this kind of contextual drift?²¹ One tempting idea would be to pair each word in a sentence with its own context. This approach generalizes from the fact that a context-sensitive expression may have different truth conditional contributions in different sentences due to a difference in extralinguistic context. Consider, for instance, sequential utterances of (6) and (7).

(6) That is red.

(7) That is not red.

These utterances may both be true provided that the first sentence is uttered in a context c in which a red thing is demonstrated and the second sentence is uttered

²¹ See Radulescu (2015) for a development of a semantic framework that ventures in this direction, though he is concerned with context change within an *argument*, not a sentence.

in a context c^* in which a non-red thing is demonstrated. The standard Kaplanian theory of demonstratives can accommodate this, since it evaluates each sentence against a distinct context—(6) is true relative to context c where it is uttered and (7) is true relative to context c^* where it is uttered. Thus, Kaplan's theory has no problem with occurrences of the same demonstrative *in different sentences* referring to different objects—sentences can be evaluated at different contexts. But when we turn to a single sentence our hands are tied—every part of a sentence must be evaluated at the same context. Thus, the two occurrences of 'that' in (8) are forced to refer to the same object.

(8) That is red and that is not red.

It seems that we must relax the Kaplanian constraint, so that the parts of a sentence can be evaluated at different contexts. The 'one-context-per-word' strategy would first resolve all context-sensitivity by pairing each word with a context, and would then compose the resulting values to yield a truth condition. Since each word can be evaluated at a distinct context, there will be no problem with recurring demonstratives or indexicals. If the first 'that' is paired with c and the second 'that' is paired with c^* , and c and c^* have distinct salient demonstrata, then it could be that $\llbracket \text{that} \rrbracket^c$ is red but $\llbracket \text{that} \rrbracket^{c^*}$ isn't.

Although initially tempting, we think this idea is misguided. The input to composition are atomic expressions α paired with contexts c , which we can abbreviate as α^c , so instead of writing $\llbracket \text{that} \rrbracket^c$ we might as well write $\llbracket \text{that}^c \rrbracket$ to emphasize this fact.²² Thus we get a clause for 'that' such as

$\llbracket \text{that}^c \rrbracket = \text{the salient demonstratum in } c.$

This provides the character for 'that', but we are lacking a definition of the character of complex expressions, including the character of sentences. One might take the discussion above to suggest the following:

$\llbracket \text{that}^c = \text{that}^{c^*} \rrbracket = 1$ iff the salient demonstratum in $c = \text{the salient demonstratum in } c^*$

But this doesn't give a definition of the character of 'that is identical to that'—it doesn't provide the sentence's profile across contexts. And if there are no semantic rules for evaluating a sentence in a context, then there is no account of utterance truth.²³ For

²² Putting it this way makes it look a lot like the Kaplanian view whereby extra-linguistic reality completes an expression—the input to semantic evaluation is a hybrid consisting of the word and a context (or a demonstratum). And since we can just think of a context here as a demonstratum, the basic meaningful units are (demonstrative, demonstratum) pairs; or if we are talking about the personal pronoun, then it is pairs of 'I' and a speaker. Thus, that^c and that^{c^*} appear to be distinct 'lexical' items.

²³ Likewise, Recanati (2010: 85–7) denies that the linguistic meaning of a demonstrative (or expression generally) determines its truth-conditional content in a context—instead the linguistic meaning plus a context merely *constrain* the truth-conditional content. This is to deny (D2). The truth-conditions are reached only by a pragmatic matching process (which appeals to uses of expressions and 'speaker's meaning'): 'a demonstrative refers to what the speaker who uses it refers to by using it.' We concede

an utterance of a sentence ϕ to be true, it must be true relative to some context c . We require a clause such as the following:

ϕ is true in c iff . . .

But on the proposal under consideration, there just is no such story, and it is unclear how it might be told. There is no story about how the contexts in which each ‘that’ occurs (i.e. c and c^*) relate to the context in which the whole sentence is uttered.²⁴

$\llbracket \text{that}^c = \text{that}^{c^*} \rrbracket = \llbracket \text{that} = \text{that} \rrbracket^{(?)}$

Context is not a parameter against which sentences (or complex expressions generally) are even evaluated. Instead each expression occurs in its own context, and the resulting contents of those expressions in their various contexts determine the ultimate truth conditions of the sentence in its—unspecified but presumably somehow related—context of utterance.

Thus, without supplementation this view has no account of utterance truth.²⁵ Given that it has no account of utterance truth, it cannot vindicate the platitude that the sentence ‘That is identical to that’ is false in some contexts. The view, thereby, also has little hope of offering a *logic* of demonstratives, which seemingly must be defined in terms of preservation of truth at a context. As a result of these difficulties, we don’t think the correct account should appeal to uncontrolled contextual drift. Instead the correct account must describe the context as evolving in accordance with semantic rules from an input context to an output context in the processing of a sentence. We call this, controlled context shifting.

5.3.2 Braun’s controlled context shifting

The previous section showed that if each utterance of a demonstrative in a sentence is to be evaluated at a different context, then the evolution of the context must be

that *anaphora* resolution may require this pragmatic process. However, we aim at a systematic account of *demonstrative* uses of pronouns, since such an account is necessary if we want anything like a *logic of demonstratives*.

²⁴ Radulescu (2012: 86) makes this point nicely as follows:

One idea is to allow one context per word; this seems to me the most promising option, but there are technical difficulties which demand further attention. (For instance, truth is normally defined as truth in a context; if we have several contexts per sentence, with respect to which of those should we evaluate the sentence? Surely not all, since the facts may change from one context to another; but then, which one?)

²⁵ One might try to supplement the proposal by defining the truth of a sentence relative to a *sequence of contexts*: $\llbracket \text{that}^{c_1} = \text{that}^{c_2} \rrbracket = \llbracket \text{that} = \text{that} \rrbracket^{(c_1, c_2)}$. But this doesn’t really help, since we now need to be told how to go from a sequence of contexts to a truth conditional contribution. How is the following defined: $\llbracket \text{that} \rrbracket^{(c_1, \dots, c_n)}$? More generally, for any sentence ϕ , and contexts c_1, \dots, c_n , we would need a definition of $\llbracket \phi \rrbracket^{(c_1, \dots, c_n)}$ in terms of the truth conditional contributions at (c_1, \dots, c_n) of the parts of ϕ . But how? A related alternative would be evaluate a complex expression at a context in terms of its parts at sub-contexts, where sub-contexts bear something like a parthood relation to the original context (see Recanati 2010, 44–5 for a suggestion along these lines).

governed by semantic rules as the sentence is processed. Otherwise, utterance truth will not be well defined as truth at the context of utterance. If there is controlled context shifting, then the truth conditions of a sentence ϕ evaluated at c may depend on the linguistic meanings of its constituent expressions evaluated at a different context c^* .

Kaplan (1989a) famously observes that in a sentence such as ‘In some contexts, I am hungry’, the content of ‘I’ does not shift under the operator. He insists that the operator ‘In some contexts...’ cannot shift the context against which one evaluates the embedded sentence. And more generally, he posits that there are no such operators—such devices would be semantic *monsters*.²⁶

Operators like ‘In some contexts it is true that’, which attempt to meddle with character, I call *monsters*. I claim that none can be expressed in English . . . And such operators could not be added to it. (Kaplan 1989a: 520–1)

This supposition provides an elegant account of the distinct roles of context and circumstances of evaluation. Principle (D3) of the antinomy is equivalent to Kaplan’s prohibition against monsters.²⁷ A monstrous operator shifts the context and thereby can shift the contribution that an expression makes to the truth-conditions of a sentence (in a context c) that contains it, whereas (D3) mandates that the truth-conditions of a sentence in a context c depend on the truth-conditional contributions of its constituents relative to c . To resolve the antinomy by rejecting (D3), one must find a context-shifting operator in the relevant sentences. In the sentence at issue—‘that is identical to that’, the only relevant candidate is the demonstrative ‘that’ itself. Thus, one who resolves the antinomy by rejecting (D3) must posit that demonstratives *themselves* are monsters. This is the proposal of Braun (1996) and a part of our own proposal.

Braun (1996) develops the most sophisticated theory to date of the evolution of context in processing a sentence with multiple demonstratives. Actually, he develops two theories. We will focus on the simpler one, but what we say may be carried over to Braun’s preferred, more sophisticated theory. Braun’s theory begins by offering a richer characterization of extralinguistic context than has been offered so far. He includes in each context c a sequence of demonstrated individuals, d_c , where the i th demonstrated individual is $d_{i,c}$. Of course, including such a sequence of demonstrated individuals in a context is not sufficient to assign a demonstrative pronoun to its referent, since the same demonstrative may occur twice in a sentence as in ‘that is identical to that’.

²⁶ It seems that for Kaplan it is not merely a contingent fact about English that such monstrous operations do not exist, but rather it is a deep, central, non-contingent fact grounded in the nature of context-sensitivity, linguistic content, and the principle of compositionality. This is why Kaplan constantly insists that we must sharply distinguish the roles of context and circumstance. The role of context is to *generate* content, while the role of circumstance is to *evaluate* content. See Rabern and Ball (forthcoming) for discussion.

²⁷ See Rabern (2013) and Rabern and Ball (forthcoming) for discussion of Kaplan’s monster prohibition in relation to the compositionality of content in a context.

Thus, Braun's contexts also contain a privileged individual, which he calls the *focal demonstratum*. He labels the focal demonstratum in c as d_c^* . Thus, Braun (1996: 165) has made two crucial assumptions about each context, which we quote verbatim:

- (a) For every context c , there is exactly one denumerable sequence of individuals which is *the sequence of demonstrata in c* (in symbols, d_c). The i -th member of this sequence is the i -th demonstratum of c (in symbols, $d_{i,c}$).
- (b) For every context c , there is exactly one member of d_c , which is *the focal demonstratum of c* (in symbols, d_c^*).

These two features account for the semantic function of a demonstrative. In particular, an utterance of a demonstrative in a context c has two effects. First, it refers to d_c^* , the focal demonstratum in c . Second, it identifies the focal demonstratum in the salience ranking as $d_{i,c}$ and outputs a new context in which the next individual in the salience ranking is now the focal demonstratum.²⁸

In Braun's (1996: 165-6) explicit semantics, an occurrence of expression α in a context c will result in a new context, $c^+ = c[\![\alpha]\!]$. The linguistic meaning of a demonstrative, 'that', in a context c determines both a referent $[\text{that}](c)$, and a context shift $c[\![\text{that}]\!]$. Thus, Braun offers the following two-part semantic characterization of 'that':²⁹

Reference: $[\text{that}](c) = d_c^*$

Shift: If $d_c^* = d_{i,c}$, then $c[\![\text{that}]\!] = c^+$ such that c^+ agrees with c except that $d_c^* = d_{i+1,c}$

An atomic sentence such as 'that is identical to that' is true at a context c just in case the focal demonstratum of the context and its successor focal demonstratum are identical. More generally, the truth conditions for atomic sentences can be specified as follows:

$$[F\alpha_1\alpha_2 \dots \alpha_n](c) = 1 \text{ iff } \langle [\alpha_1](c), [\alpha_2](c[\![\alpha_1]\!]), \dots, [\alpha_n](c[\![\alpha_1]\!][\![\alpha_2]\!] \dots [\![\alpha_{n-1}]\!]) \rangle \in I(F)$$

The sentence 'that is identical to that' can be false at a context c if the focal demonstratum at c is not identical to the focal demonstratum at the successor context, c^+ .

This is on the right track. Unfortunately, there is a glitch. Braun's context-shifting theory attempts to model the shift solely by appeal to *extralinguistic* context—the object demonstrated—but it thereby lacks the required resources to account for the update. In particular, the only resources in terms of which Braun describes the shift in context induce by a demonstrative are (a) the sequence of demonstrata and (b) the focal demonstratum (the object itself). This is insufficient, for multiple occurrences of a demonstrative 'that' in a sentence may refer to the same individual, as happens

²⁸ In Braun's more sophisticated theory, the sequence of demonstrata is replaced by a sequence of characters and the focal demonstratum is replaced by the *operative character*.

²⁹ Note that for clarity of presentation we make a few alterations to Braun. We use the postfix notion ' $c[\![\alpha]\!]$ ' for Braun's shift function notation ' $s(\alpha, c)$ '. And we are offering a truth conditional semantics, where Braun offers a semantics in terms of structured propositions.

in any *true* utterance of ‘that is identical to that’. This can be made most explicit by considering a sentence with multiple occurrences of a demonstrative such as (9).

(9) *That* is identical to *that*, but not *that*.

Sentence (9) is true provided that the first occurrence of ‘that’ refers to the same object as the second occurrence of ‘that’, but a different object from the third occurrence of ‘that’. Thus, it may be true if the first and second occurrences of ‘that’ refer to the same hat, but the third occurrence refers to a book. Braun’s theory should have the resources to predict that this sentence is true in some contexts. But his theory is unable to do so.

According to Braun’s theory, sentence (9) is true in c just in case the focal demonstratum in c is identical to the focal demonstratum in the successor context of c , but not identical to the focal demonstratum of the following context. Context c will be modelled as a sequence of individuals, say $d_c = \langle a, a, b \rangle$, paired with a focal demonstratum, $d_c^* = a$. To assess (9) for truth in c we must be able to assess the result of shifting c to the successor context. But herein lies the problem. The shift is defined in terms of where the focal demonstratum $d_c^* = a$ occurs in the sequence of demonstrata $d_c = \langle a, a, b \rangle$. (Recall: if $d_c^* = d_{i,c}$, then $c[\text{that}] = c^+$ such that c^+ agrees with c except that $d_c^* = d_{i+1,c}$.) But, of course, a occurs more than once in $\langle a, a, b \rangle$. Thus, the shift function is not well defined for this sequence. But there’s nothing special about this sequence. The problem emerges whenever the same object is available for reference in multiple positions.

It’s worth also putting this objection in the following alternative way: Braun’s context update function is defined in terms of the referent of a demonstrative (in context). So once we fix the initial context, if we have any two co-referential demonstrative uses, they must produce the same output context. But then getting the right result for (9) is just impossible. The first identity requires that the first two demonstrative uses be coreferential, so they have to have the same update effects. Since updating with the first left the focal demonstratum unchanged, so will the second, so we can’t get the third demonstrative non-co-referential with the first two. Since Braun’s theory is precisely designed to deliver the result that multiple occurrences of a demonstrative in a sentence sometimes co-refer and sometimes don’t, his theory is inadequate.³⁰

A further inadequacy in Braun’s view is that it can’t be extended to anaphora. What good is a univocal treatment of demonstrative pronouns, if we need them to be

³⁰ The difference between Braun’s view and our preferred view might appear merely technical. Of course, in a sense it is, but we highlight the glitch since we think that the technical difference actually manifests an important philosophical difference. Braun’s implementation of the update is motivated by background commitments to ‘direct reference’. But once we add in the requisite sensitivity to discourse context the semantic contribution of a demonstrative evaluated at a (local) context cannot be characterized solely in terms of its referent (at the context). Thus this is a compromise of direct reference, given that the semantics proceeds at a layer of representation intermediate between the demonstrative and its referent. (Note that the technical problem for the context shifting theory applies to Braun’s more sophisticated character-shifting theory, so long as the same character can occur twice in the sequence of characters. And, Braun’s explicit goal is to allow for this possibility.)

ambiguous anyhow to serve as antecedents of distinct anaphoric pronouns? By way of contrast, the view we develop in the next section can naturally be extended to anaphora. We should mention in this connection that Georgi (2015) offers a development of Braun's theory that is structurally similar to our own, though his reasons for departing from Braun differ from ours. Georgi seems to think Braun's theory is formally adequate, but too philosophically committed. In particular, Georgi wants to be neutral as to whether his theory actually mandates a shift in *context* as opposed to some other semantic input (Georgi 2015: §2.1 and footnote 18). Georgi's theory—like Braun's—is subject to our second criticism insofar as it does not (and is not intended to) handle anaphora. But this lacuna reintroduces the threat that demonstratives will need to be ambiguous in order to serve as antecedents of anaphoric pronouns.³¹

5.4 Recurring Demonstratives and Discourse Context

The discussion above led to two results. First, the sentence 'that is identical to that' is univocal as are both occurrences of the demonstrative 'that'. Second, as we saw in the discussion of context-shifting strategies, the difference in the truth conditional contributions of two occurrences of a demonstrative cannot arise solely from a difference in extralinguistic context. In particular, it's not enough to treat the demonstrative as shifting the object demonstrated (or the demonstration itself), because the same object (or demonstration) can recur in an extralinguistic context at which we evaluate a discourse.

We solve this problem with the context-shifting strategy by supposing that what an occurrence of a demonstrative shifts is not the object demonstrated (or demonstration), but rather the index at which one evaluates the next demonstrative. That is, rather than treating a demonstrative as evaluable at a sequence of demonstrata and a focal demonstratum that happens to occur in the sequence, we should treat the demonstrative as semantically evaluable at a sequence and a numerical index, which determines a position in the sequence. So on our view—in contrast to Braun's—the context at which a demonstrative is evaluated includes information about prior discourse. To put this another way, referential indices are elements of *discourse* context and not *extralinguistic* context. On our proposal, a demonstrative can only be

³¹ Georgi does provide a coordination schema as an input to the semantics which is meant to characterize *de jure* co-reference. The coordination scheme r is an equivalence relation on positions in the sequence of demonstrata σ . Essentially it imposes a restriction on the sequence such that if $r(i, j)$, then the same object occurs at both the i^{th} and j^{th} position of σ . In a case where the coreference is non-accidental, the two occurrences of 'that' would be co-referential relative to any context that shares the coordination scheme—whereas in the accidental case this does not hold. In this way Georgi can get something like 'anaphora'. But it would be misguided to appeal to this mechanism to account for anaphoric relations in general. Such an account would essentially treat all cases of anaphoric dependence of one pronoun on another as just co-referential pronouns. In other words, it would characterize anaphora simply by placing constraints on the sequence of demonstrata at which the sentence (or discourse) is evaluated. For this reason, there would be no prospect for extending this account to pronouns which are anaphoric on quantified noun phrases.

evaluated with respect to both parameters. Moreover, each occurrence of a demonstrative in a discourse has the effect of *shifting* the discourse context, by shifting the index against which the next demonstrative is evaluated. In this respect and in this respect only, each occurrence of a demonstrative in a sentence is evaluated at a distinct context.

One might worry that this proposal involves a return to the syntactic indexing strategy. But it does not. The syntactic indexing strategy posited that demonstrative pronouns are *ambiguous*. On our view, demonstratives are univocal, but context sensitive. Specifically, one semantically processes a sentence containing a demonstrative pronoun by evaluating the pronoun at the index provided by prior discourse context and the extralinguistic context provided by the utterance. Thus, ‘referential indices’ are not *syntactic objects*, but are among the points of evaluation.³²

Building on the resolution to the antinomy of the variable in terms of *dynamic indexing* (see Pickel and Rabern 2016), our theory extends the account to treat the indices on demonstratives as purely semantic objects—in the sense that they are parameters that arise out of semantic processing. The discourse context begins without any referential indices. Each occurrence of a demonstrative introduces a new index and refers to the corresponding object in the salience ranking provided by extralinguistic context. Treating demonstratives as shifting discourse context primes us to offer a unified account of demonstrative and anaphoric pronouns, since the latter are obviously sensitive to discourse context. Moreover, it offers the hope of a comprehensive account of how noun phrases receive a referential index in general.³³

According to our view, an occurrence of ‘that’ shifts the discourse context. To model this, we need to explicitly distinguish extralinguistic context from discourse context. Let *c* denote the standard extralinguistic context, construed as a sequence of salient individuals. Let *d* denote discourse context, construed as a set of pairs of noun phrases from the prior discourse and their referential indices. (The lexical item serves as a mere placeholder for its ϕ -features, instead of $\{(\text{that}, 1)\}$, the final theory might have $\{((3rd, neuter, singular), 1)\}$. These ϕ -features may, in a fuller treatment, act as a definedness constraint on updates and denotations.³⁴) Then we can model

³² It’s not original to think of referential indices as entering into *semantic* processing rather than having purely syntactic effects. In Heim’s (1982: §5.1) theory, each new sentence is evaluated for felicity against a prior set of referential indices and outputs a new set of discourse referents, which she identifies with referential indices (Heim 1982: 165). This means that referential indices are, at least in part, *semantic objects*. (Yet, the novelty-familiarity condition, which distinguishes definites and indefinites can also be viewed as, at least in part, a syntactic constraint, cf. Yalcin 2012. Our picture promises a natural and wholly semantic explanation of such phenomena.)

³³ Treating referential indices as semantic objects in no way conflicts with binding constraints on coindexing (Heim and Kratzer 1998: §5.5). These constraints mandate, among other things, that certain bound pronouns c-commanded by their antecedents must be made reflexive.

³⁴ First and second person pronouns could be assimilated to the demonstrative case with appropriate ϕ -features, and possibly names as well. Phi theory is underdeveloped and issues such as the inventory of features (person, number, gender, etc.) and their exact syntactic and semantic status remain unsettled (see Harbour et al. 2008).

the discourse effect of an occurrence of a demonstrative as adding to the discourse context the pair consisting of the demonstrative type and a new referential index. The referential index assigned will be the successor to the highest index assigned by the prior discourse d , which we denote by $\text{MAX}(d)$. Thus, we add $\text{MAX}(d) + 1$. More explicitly, the update rule for ‘that’ can be specified as follows:

$$d[\text{that}] = d \cup \{(\text{that}, \text{MAX}(d) + 1)\}$$

Thus, in processing the sentence ‘that is identical to that’, we might begin in the null discourse context $d = \emptyset$. The first occurrence of ‘that’ will shift the discourse context to $d[\text{that}] = \{(\text{that}, 1)\}$. The second occurrence of ‘that’ will then shift the discourse context again to $d[\text{that}][\text{that}] = \{(\text{that}, 1), (\text{that}, 2)\}$. The occurrences of ‘that’ are, thereby, evaluated at different discourse contexts which enables them to make different truth conditional contributions despite having the same linguistic meaning.

The truth conditional contribution of a demonstrative in an extralinguistic context at a particular point in discourse can be extracted from its update. In particular, in a discourse context d and extralinguistic context c , ‘that’ refers to the object in the position in the salience ranking corresponding to the next available referential index. We define a function $[\alpha](c, d)$ that maps a context to the referent of α relative to that context.

$$[\text{that}](c, d) = c_{\text{MAX}(d^*)}, \text{ where } d^* = d[\text{that}]$$

If multiple pronouns are used in a sentence, the first pronoun corresponds to the first position in the salience ranking, the second pronoun corresponds to the second position in the salience ranking, and so on. More generally, each occurrence of ‘that’ shifts the discourse context by adding one referential index. This means that the i^{th} occurrence of the demonstrative, ‘that’, in a sentence evaluated at context (c, d) will refer to the following:

$$[\text{that}](c, d \overbrace{[\text{that}][\text{that}] \dots [\text{that}]}^{(i-1)\text{-times}})$$

We can now define the truth conditions of an atomic sentence predicating the n -ary relation F^n of n occurrences of ‘that’: ‘ $F^n(\text{that}, \dots, \text{that})$ ’. As the atomic sentence is processed, the first demonstrative is assessed relative to the first context, the second demonstrative is assessed relative to the second context, and so on. The sentence is true in a context (c, d) just in case F^n relates the referent of the first demonstrative at its relevant context to the second demonstrative at its relevant context and so on. More formally, one may say (where I is the interpretation function that maps an n -ary predicate to a set of n -tuples drawn from the domain):

$$[F^n(\text{that}, \dots, \text{that})](c, d) = 1 \text{ iff } \langle [\text{that}](c, d), \dots, [\text{that}](c, d \overbrace{[\text{that}][\text{that}] \dots [\text{that}]}^{(n-1)\text{-times}}) \rangle \in I(F^n)$$

This truth condition mimics the truth condition that a subscript strategy might assign. But it does so compositionally and without positing that the demonstrative ‘that’ is ambiguous.

The semantics offered so far provides only a truth condition for an atomic sentence. But, in order to assess complex sentences such as conjunctions and conditionals, we will also need to posit an update condition for each sentence. Effectively, an atomic sentence will update the discourse context by adding the referential indices of all of its demonstrative pronouns. A conjunction will update the discourse context by first updating with the referential indices of its first conjunct and then updating with the referential indices of its second conjunct. (See the appendix for the complete semantics.) Such a dual specification of truth (or satisfaction) conditions and update conditions mimics Heim’s (1982) file change semantics in separating the update and truth conditional effects. However, this is entirely optional, since the truth conditional effects can be folded into the update condition just as Groenendijk and Stokhof (1991) do for their development of Heim’s semantics.³⁵

At this point, we have solved the problem of recurring demonstratives by rejecting the claim that each demonstrative in a sentence is assessed relative to the same context. The sentence ‘that is identical to that’ may be false at a context (c, d) , where $d = \emptyset$ and $c = \langle a, b \rangle$ such that $a \neq b$. This can be seen by evaluating the problem sentence at this context in accordance with our semantic clauses:

$$\begin{aligned} [\mathbf{that} = \mathbf{that}]((\langle a, b \rangle, \emptyset)) &= 1 \text{ iff} \\ [\mathbf{that}]((\langle a, b \rangle, \emptyset)) &= [\mathbf{that}]((\langle a, b \rangle, \emptyset \llbracket \mathbf{that} \rrbracket)) \text{ iff} \\ [\mathbf{that}]((\langle a, b \rangle, \emptyset)) &= [\mathbf{that}]((\langle a, b \rangle, \{(\mathbf{that}, 1)\})) \text{ iff} \\ a &= b \end{aligned}$$

Since by assumption $a \neq b$ we get the desired result that $[\mathbf{that} = \mathbf{that}]((\langle a, b \rangle, \emptyset)) \neq 1$ (i.e. sentence (3) is false at some contexts). Our solution essentially involves incorporating elements of prior discourse into context and specifying systematic semantic rules for the evolution of discourse context. In particular, our resolution of the problem of recurring demonstratives comes through replacement of (D3) with the claim that the truth-conditions of a sentence in a context are determined by the truth-conditional contributions of the occurrences of its constituent expressions relative to *suitable updates of the context* and their modes of combination. It is superior to prior context-shifting strategies because it is systematic, and it explains why discourse shifts

³⁵ Note that this proposal is also compatible with structured content approaches such as those adopted by Braun (1996). Rather than thinking of each sentence as expressing a structured content, one should think of a discourse as expressing structured content. Dynamic construction algorithms specify how sentences update this structured content. The idea that a discourse as a whole rather than its individual sentences expresses a structured content might be inspired by Kamp (1981), according to whom each discourse expresses a discourse representation structure.

the context. It is also superior to syntactic indexing or linking strategies because it maintains the univocality of demonstratives.

But our account promises more. The framework can be generalized to explain the evolution and semantic effects of the discourse referents (that is, indices) associated with *all* noun phrases in a discourse. This would provide an account of anaphoric pronouns which does not require indexing the antecedents of anaphoric pronouns. Without an account, we have independent reason to reject the univocality of uses of demonstratives. We develop such an account in the following section.

5.5 Anaphora

Our aim now is to provide a semantics for anaphoric pronouns that does not require ambiguity in the antecedents to anaphoric pronouns. This is especially relevant because *demonstratives* themselves can be antecedents of anaphoric pronouns. Our semantics for demonstrative pronouns introduced a rich discourse structure which evolves as further occurrences of demonstratives are introduced. We will now develop a semantics for anaphoric pronouns which exploits this rich structure.

The semantics we develop makes use of two simplifications that we want to flag up-front, because they are the last vestiges of ambiguity required by our semantics for demonstrative and anaphoric pronouns.

OVERSIMPLIFICATIONS:

- Demonstrative and anaphoric pronouns are syntactically differentiated by appending an anaphoricity operator \uparrow to the anaphoric pronoun.
- The operator appended to an occurrence of an anaphoric pronoun will bear an index: \uparrow_i . However, the antecedents of anaphoric pronouns remain univocal.

Both oversimplifications are ultimately unnecessary. Anaphoric and demonstrative pronouns can be given a uniform semantic treatment at the cost of some under-specification.³⁶ However, for the purposes of this chapter we adopt the more limited aim of showing that we can give a semantics for anaphoric pronouns which does not require indices on their antecedents. This will take us closer to completing the parameterization revolution.

To be clear, we aim only to show that the antecedents of anaphoric pronouns can be univocal. Thus, on our approach (2a) will be univocal while (2b) will have different interpretations.

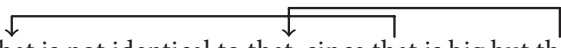
- (2) (a) An engineer saw an engineer.
 (b) An engineer saw an engineer and she waved.

³⁶ When we come to incorporate anaphora, we will slightly relax (D2), in that one will first need to resolve whether a pronoun is deictic or anaphoric. However, given a deictic use of a pronoun, our account preserves (D2).

This is a desirable result, since there is genuine unclarity in identifying the antecedent of ‘she’ in (2b). This unclarity requires further complexity. Nonetheless, there is no corresponding unclarity in (2a). The task then is to explain why a sequence such as (2b) has multiple resolutions while (2a) does not.

We begin by complicating our notion of a discourse context. Anaphoric pronouns, like demonstrative pronouns, will introduce a new position in a discourse context. But, unlike demonstrative pronouns, these positions must be tied or linked to their antecedents. Thus, rather than associating each expression type with a single position in a sequence, a discourse context will now associate each expression type with a *sequence* of positions. When a discourse context associates an expression with a sequence, it serves as a stack giving the ‘anaphoric history’ of the determiner phrase, showing how a sequence of discourse referent introductions have been anaphorically linked to one another as the discourse has developed.³⁷

Sentence (10) illustrates the intended evolution of a discourse context on this new conception.

- (10)  That is not identical to that, since that is big but that is not.

In (10), the first and second occurrences of ‘that’ are used deictically. The third occurrence is anaphoric on the first occurrence. The final occurrence is anaphoric on the second occurrence. We use arrows merely to indicate anaphoric dependence, and we leave open for now whether this dependence is ultimately to be represented in the syntax. In particular, we are not presupposing the syntactic analysis of anaphora relations that we rejected above. On our view, the discourse context evolves as each occurrence of the pronoun ‘that’ is processed in (10). This evolution can be represented by Figure (5.1).

The intended interpretation here is that two non-anaphoric discourse referents are introduced (marked with 1 and 2). Another discourse referent—3—is introduced as

INPUT DISCOURSE CONTEXT	TYPE OF UPDATE	OUTPUT DISCOURSE CONTEXT
{ }	that	{(that, ⟨1⟩) }
{(that, ⟨1⟩) }	that	{(that, ⟨1⟩), (that, ⟨2⟩) }
{(that, ⟨1⟩), (that, ⟨2⟩) }	that+anaphoricity	{(that, ⟨1, 3⟩), (that, ⟨2⟩) }
{(that, ⟨1, 3⟩), (that, ⟨2⟩) }	that+anaphoricity	{(that, ⟨1, 3⟩), (that, ⟨2, 4⟩) }

Figure 5.1. Evolution of a discourse context

³⁷ Our approach is related to Vermeulen (2000) who models the dynamics of variables in first-order logic using stacks of *individuals* in the domain of quantification. Haug (2014) modifies this approach in the context of Discourse Representation Theory by modeling variables in terms of stacks of *registers* of these individuals.

anaphorically linked to 1, before a further discourse referent 4 is introduced and linked to 2.

To model this evolution of discourse context, we divide the update induced by an anaphoric pronoun in two. First, the pronoun ‘that’ updates the context as before by introducing a new discourse referent $\langle \mathbf{that}, \text{MAX}(d) + 1 \rangle$. The anaphoricity operator \uparrow_n , then further updates the context by merging the index introduced by the occurrence of ‘that’ with some previously introduced index. In particular, this update ‘fuses’ the most recent discourse referent in d (containing $\text{MAX}(d)$) with the anaphoric chain associated with the demonstrative occurrence n points earlier in the discourse. It thereby serves to link together discourse referents already in place in the discourse context into an anaphoric chain. (The formal details are presented in the appendix.)

We capture the anaphoric reading of ‘that is identical to that’ with the syntax ‘that is identical to that \uparrow_1 ’, where the role of the anaphoric marker \uparrow_1 is to indicate that the pronoun ‘that’ is to be merged with the most recent anaphoric chain. Our syntactic representations of anaphora—whereby the antecedents are not distinguished with indices but the anaphoric pronouns indicate how far back their antecedent is—has a strong affinity to the use of so-called ‘De Bruijn indices’ to represent binding relations in the lambda calculus. De Bruijn (1972) developed a notional device for the lambda calculus specifically designed to overcome problems stemming from alpha equivalence and unwanted variable capture. In this notation alphabetic variants get the same representation since the occurrence of a bound variable is replaced with a natural number indicating the ‘distance’ to its antecedent. For example, $\lambda x.\lambda y.x$ and $\lambda x.\lambda y.y$ are rendered as $\lambda\lambda 2$ and $\lambda\lambda 1$, respectively. This notational system is put forward as an alternative way to write lambda terms in order to facilitate manipulations, thus, although there could be, there need be no semantic component *per se*.

Our picture, of course, has a semantic element as well. The anaphoric pronoun ‘that \uparrow_1 ’ updates the discourse context in two stages. First ‘that’ updates the discourse context as a pronoun followed by the anaphoricity marker ‘ \uparrow_1 ’. Thus, $d[\llbracket \mathbf{that} \uparrow_1 \rrbracket] = d[\llbracket \mathbf{that} \rrbracket][\uparrow_1]$. The evolution of the discourse context in example (10) ‘That is not identical to that, since that \uparrow_2 is big but that \uparrow_2 is not’ can now be broken down further as follows.

By sequentially updating with ‘that’ and then the indexed anaphoricity operator \uparrow_n , the semantics yields the evolution of context as characterized in Figure 5.1.

We have described how demonstrative and anaphoric uses of ‘that’ update the discourse context. We have shown how this enables different occurrences of a demonstrative to make different truth conditional contributions. What remains is to explain the mechanism by which the truth conditional contribution of an anaphoric pronoun *depends* on the truth conditional contribution of its antecedent. Here we focus on cases in which the antecedent is *referential*, though our view can be generalized to quantificational antecedents.

If the antecedent of an anaphoric pronoun is referential, then the two pronouns should make the same truth conditional contribution. That is, they should refer to

the same thing. So the referent of a demonstrative should be the same as that of any pronoun anaphoric on it. In our prior discussion, a demonstrative pronoun ‘that’ affects the discourse context d by adding a new pair $\langle \mathbf{that}, \text{MAX}(d) + 1 \rangle$. The truth conditional contribution of the demonstrative was given as the $(\text{MAX}(d) + 1)^{\text{st}}$ position in the extra-linguistic context c , or $c_{\text{MAX}(d)+1}$. So the truth conditional contribution of the n^{th} introduced demonstrative was given by the n^{th} position in the extra-linguistic context.

The present story is not quite so simple since we are dealing with anaphoric chains generated by multiple occurrences of pronouns rather than single pronouns. We still want to say that the n^{th} anaphoric chain introduced is associated with the n^{th} position in a sequence. We do so, by ordering anaphoric chains by their first introduction in the discourse. We then give the truth conditional contribution of a pronoun which contributes the n^{th} introduced anaphoric chain as the n^{th} position in the extra-linguistic context. In the case of a demonstrative pronoun, very little changes. In the case of an anaphoric pronoun, we identify its truth conditional contribution with that of its original antecedent.

Consider again sentence (10) above.

- (10) That is not identical to that, since that \uparrow_2 is big but that \uparrow_2 is not.

Processing this sentence resulted in the discourse context $d = \{\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2, 4 \rangle \rangle\}$ (Figure 5.2). Here we have two anaphoric chains. The first and third occurrences of ‘that’ are linked to the first-introduced anaphoric chain, $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle$, and thus will refer to the first position in the extra-linguistic context. The second and fourth occurrences of ‘that’ are linked to the second-introduced anaphoric chain, $\langle \mathbf{that}, \langle 2, 4 \rangle \rangle$, and thus will refer to the second position in the extra-linguistic context. The sentence will, therefore, be true just in case the first demonstratum is not identical to the second demonstratum as evidenced by the fact the former is big, but the latter isn’t. This conforms to our judgment of the truth conditions of the sentence.

INPUT DISCOURSE CONTEXT	UPDATE	OUTPUT DISCOURSE CONTEXT
{	$\llbracket \mathbf{that} \rrbracket$	{ $\langle \mathbf{that}, \langle 1 \rangle \rangle$ }
{ $\langle \mathbf{that}, \langle 1 \rangle \rangle$ }	$\llbracket \mathbf{that} \rrbracket$	{ $\langle \mathbf{that}, \langle 1 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle$ }
{ $\langle \mathbf{that}, \langle 1 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle$ }	$\llbracket \mathbf{that} \rrbracket$	{ $\langle \mathbf{that}, \langle 1 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle, \langle \mathbf{that}, \langle 3 \rangle \rangle$ }
{ $\langle \mathbf{that}, \langle 1 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle, \langle \mathbf{that}, \langle 3 \rangle \rangle$ }	$\llbracket \uparrow_2 \rrbracket$	{ $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle$ }
{ $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle$ }	$\llbracket \mathbf{that} \rrbracket$	{ $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle, \langle \mathbf{that}, \langle 4 \rangle \rangle$ }
{ $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2 \rangle \rangle, \langle \mathbf{that}, \langle 4 \rangle \rangle$ }	$\llbracket \uparrow_2 \rrbracket$	{ $\langle \mathbf{that}, \langle 1, 3 \rangle \rangle, \langle \mathbf{that}, \langle 2, 4 \rangle \rangle$ }

Figure 5.2. Dynamics of a discourse context with anaphora

5.6 Conclusion

The parameterization revolution promised to explain how a univocal expression could make distinct truth conditional contributions in its various occurrences. But it stalled on account of the problem of recurring demonstratives and also on account of the need to link anaphoric pronouns to their antecedents. Semanticists have been too content to posit massive ambiguities in demonstrative pronouns. They have been consoled by the thought that this ambiguity would ultimately be needed anyhow to explain anaphora. We have revived the spirit of the revolution by showing how to treat demonstrative pronouns as univocal and providing an account of anaphora that doesn't end up re-introducing the ambiguity.

In the case of recurring demonstratives, we have argued that the discourse context evolves as various occurrences of a demonstrative are processed. The truth conditional contributions of a demonstrative varies both with the extra-linguistic context and the discourse context at which it is processed. Therefore, the distinct occurrences of a demonstrative can have the same linguistic meaning and yet make distinct truth conditional contributions. Moreover, we have shown that our account of the context shifts induced by a demonstrative are superior to the most widely known accounts, such as that of Braun (1996).

We then extended this account to anaphoric pronouns, showing that the link between an anaphoric pronoun and its antecedent could be established without rendering the antecedent ambiguous. As we have mentioned, our semantics is oversimplified in two ways. It differentiates demonstrative and anaphoric uses of pronouns. And, it added syntactic indices to anaphoric pronouns. But the anaphoricity operator and its index supplement *only* the anaphoric pronoun and not its antecedent. The processing of the antecedent will in no way require processing of subsequent expressions anaphoric on it. Even these last vestiges of the ambiguity strategy can be overcome. But the thoroughgoing revolution must be carried out in future work.

Appendix

We have informally explained how a demonstrative or anaphoric pronoun should update the discourse context. It remains to make this discussion rigorous. We begin with the update potentials. In order to specify how a demonstrative or anaphoric pronoun updates the context, we need to rigorously characterize three notions. First, $\text{MAX}(d)$ finds the largest number of any anaphoric chain in the discourse context d .

Definition. For any discourse context d , $\text{MAX}(d)$ = the least n such that for any $\langle \epsilon, \sigma \rangle \in d$ and for any i , $n \geq \sigma_i$.

The function SEL takes a discourse context and a number and selects the unique anaphoric chain of the discourse context containing that number.

Definition. For any discourse context d and any positive integer n , $\text{SEL}(d, n) = \text{the } \langle \epsilon, \sigma \rangle \in d \text{ such that for some } i, n = \sigma_i$.

Finally, the function \oplus takes an anaphoric chain and a number and adds the number to the end of the anaphoric chain.

Definition. For any element of a discourse context $\langle \epsilon, \sigma \rangle$ where σ has length j and any positive integer n , $\langle \epsilon, \sigma \rangle \oplus n = \langle \epsilon, \langle \sigma_1, \dots, \sigma_j, n \rangle \rangle$.

Using these notions we can characterize the update potential of a demonstrative or anaphoric use of a pronoun. When any pronoun is used, it updates the discourse context by ‘advancing to the next tag.’ For example, every use of ‘that’ (whether anaphoric or deictic) will introduce a new discourse referent into the discourse context.

$$d[\llbracket \text{that} \rrbracket] = d \cup \{ \langle \text{that}, \langle \text{MAX}(d) + 1 \rangle \rangle \}$$

If a pronoun is used anaphorically, it will also merge the largest number in any anaphoric chain in the discourse context with the chain containing its antecedent determined by counting back n . Thus, the update clause for an anaphoricity operator is:

$$d[\llbracket \uparrow_n \rrbracket] = (d \setminus \{ \text{SEL}(d, \text{MAX}(d)), \text{SEL}(d, \text{MAX}(d) - n) \}) \cup \{ \text{SEL}(d, \text{MAX}(d) - n) \oplus \text{MAX}(d) \}.$$

That handles the dynamics. But in order to rigorously characterize the truth conditional contribution of a pronoun, we need a function that finds the lowest number, BOT , in any anaphoric chain, generated by the ultimate antecedent of any pronoun giving rise to the chain.

Definition. For any element of a discourse context $\langle \epsilon, \sigma \rangle$, $\text{BOT}(\langle \epsilon, \sigma \rangle) = \text{the largest } n \text{ such that for any } i, n \leq \sigma_i$.

We can now order the anaphoric chains in any discourse context by their bottom elements. A pronoun which gives rise to the number in the n^{th} anaphoric chain interacts with the n^{th} position in the extralinguistic context. We now specify the truth conditional contribution of a pronoun in an extralinguistic and discourse context (c, d) as follows. In order to express the point in maximum generality, we will take the pronouns to be of the form ‘that \uparrow_i ’ so that in the case of a demonstrative pronoun $i = 0$. (Which we use as shorthand for the *absence* of an anaphoricity operator.)

$$[\llbracket \text{that } \uparrow_i \rrbracket](c, d) = c_j,$$

where $d^* = d[\llbracket \text{that } \uparrow_i \rrbracket]$ and j is the cardinality of $\{ \langle \epsilon, \sigma \rangle \in d^* \mid \text{BOT}(\langle \epsilon, \sigma \rangle) \leq \text{BOT}(\text{SEL}(d^*, \text{MAX}(d^*))) \}$

The truth conditions of an atomic sentence containing n pronouns can be given as before:

$$[\text{F}^n(\text{that } \uparrow_{i_1}, \dots, \text{that } \uparrow_{i_n})](c, d) = 1 \text{ iff}$$

$$\langle [\llbracket \text{that } \uparrow_{i_1} \rrbracket](c, d), \dots, [\llbracket \text{that } \uparrow_{i_n} \rrbracket](c, d \overbrace{[\llbracket \text{that } \uparrow_{i_1} \rrbracket] \dots [\llbracket \text{that } \uparrow_{i_{n-1}} \rrbracket]}^{(n-1)\text{-times}}) \rangle \in I(\text{F}^n)$$

A toy language. In order to see how this all works we now provide a machine against which we can test our judgments. We provide a language containing predicates, truth-functions, demonstrative pronouns, and anaphoric pronouns. Let a model for the language be a pair $\langle D, I \rangle$,

where D is a set of individuals and I is the interpretation function that maps an n -ary predicate to a set of n -tuples drawn from D . We define the contextual updates induced by the expressions of the language and then define truth (in a model) relative to an extralinguistic context and discourse context. An extralinguistic context c is a sequence of individuals drawn from D , and a discourse context d is a set of ordered pairs of pronouns α (going proxy for their ϕ -features) paired with sequences of positive integers σ .

Lexicon:

- **that, it, \uparrow , runs, loves, not, and**

Syntax: The well-formed sentences of the language are provided by the following grammar (where each anaphoric marker is merged with some $i \in \mathbb{Z}$):

- *Terms:*
 $\alpha ::= \text{that}\uparrow_i \mid \text{it}\uparrow_i$
- *Sentences:*
 $\phi ::= \alpha \text{ runs} \mid (\alpha \text{ loves } \alpha) \mid \text{not } \phi \mid (\phi \text{ and } \phi)$

Discourse updates:

- $d[\llbracket \text{that} \rrbracket] = d \cup \{(\text{that}, \langle \text{MAX}(d) + 1 \rangle)\}$
- $d[\llbracket \text{it} \rrbracket] = d \cup \{(\text{it}, \langle \text{MAX}(d) + 1 \rangle)\}$
- $d[\llbracket \uparrow_i \rrbracket] = (d \setminus \{\text{SEL}(d, \text{MAX}(d)), \text{SEL}(d, \text{MAX}(d) - i)\}) \cup \{\text{SEL}(d, \text{MAX}(d) - i) \oplus \text{MAX}(d)\}$
- For $\gamma \in \{\text{runs, loves, not, and}\}$, $d[\llbracket \gamma \rrbracket] = d$
- For any sentence $\phi = \gamma_1 \dots \gamma_n$, $d[\llbracket \gamma_1 \dots \gamma_n \rrbracket] = d[\llbracket \gamma_1 \rrbracket] \dots [\llbracket \gamma_n \rrbracket]$

Truth and denotation:

- For a term α , $[\alpha](c, d) = c_j$, where $d^* = d[\llbracket \alpha \rrbracket]$,
 and j is the cardinality of $\{(\alpha, \sigma) \in d^* \mid \text{BOT}(\langle \alpha, \sigma \rangle) \leq \text{BOT}(\text{SEL}(d^*, \text{MAX}(d^*)))\}$
- For predicates π and terms $\alpha_1, \dots, \alpha_n$,
 $[\pi^n(\alpha_1, \dots, \alpha_n)](c, d) = 1$ iff $\langle [\alpha_1](c, d), \dots, [\alpha_n](c, d[\llbracket \alpha_1 \rrbracket] \dots [\llbracket \alpha_{n-1} \rrbracket]) \rangle \in I(\pi^n)$
- For sentence ϕ , $[\text{not } \phi](c, d) = 1$ iff $[\phi](c, d) = 0$
- For sentences ϕ and ψ , $[\phi \text{ and } \psi](c, d) = 1$ iff $[\phi](c, d) = 1$ and $[\psi](c, d[\llbracket \phi \rrbracket]) = 1$

Example sentence: ‘that \uparrow_0 loves that \uparrow_0 and it \uparrow_1 runs’

$$[\text{that}\uparrow_0 \text{ loves that}\uparrow_0 \text{ and it}\uparrow_1 \text{ runs}](\langle a, b \rangle, \emptyset) = 1$$

$$\text{iff } [\text{that}\uparrow_0 \text{ loves that}\uparrow_0](\langle a, b \rangle, \emptyset) = 1 \text{ and } [\text{it}\uparrow_1 \text{ runs}](\langle a, b \rangle, \{(\text{that}, \langle 1 \rangle), (\text{that}, \langle 2 \rangle)\}) = 1$$

$$\text{iff } \langle a, b \rangle \in I(\text{loves}) \text{ and } [\text{it}\uparrow_1 \text{ runs}](\langle a, b \rangle, \{(\text{that}, \langle 1 \rangle), (\text{that}, \langle 2 \rangle)\}) = 1$$

$$\text{iff } \langle a, b \rangle \in I(\text{loves}) \text{ and } [\text{it}\uparrow_1](\langle a, b \rangle, \{(\text{that}, \langle 1 \rangle), (\text{that}, \langle 2 \rangle)\}) \in I(\text{runs})$$

$$\text{iff } \langle a, b \rangle \in I(\text{loves}) \text{ and } c_j \in I(\text{runs}), \text{ where } c = \langle a, b \rangle \text{ and}$$

$$j = \text{the cardinality of } \{(\alpha, \sigma) \in \{(\text{that}, \langle 1 \rangle), (\text{that}, \langle 2, 3 \rangle)\} \mid \text{BOT}(\langle \alpha, \sigma \rangle) \leq \text{BOT}(\langle \text{that}, \langle 2, 3 \rangle \rangle)\}$$

$$j = 2$$

$$\text{iff } \langle a, b \rangle \in I(\text{loves}) \text{ and } b \in I(\text{runs}).$$

Acknowledgements

Some older versions of this material were presented at the 2015 Pacific APA, at the PERSP Final Conference in Barcelona, and at the Meaning Sciences Workshop in Berkeley. Thanks to those audiences for helpful feedback. Also many thanks to David Braun, Michael Glanzberg, Ephraim Glick, and various anonymous referees for detailed comments and criticism. Bryan Pickel acknowledges funding from Project FFI2016-80636-P (Spanish Government).

References

- Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy* 4(2), 159–219.
- Braun, D. (1996). Demonstratives and their linguistic meanings. *Noûs* 30(2), 145–73.
- Carnap, R. (1937/1959). *Logical Syntax of Language*. New Jersey: Littlefield, Adams, and Company, Paterson.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- De Bruijn, N. G. (1972). Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Indagationes Mathematicae* 75(5), 381–92.
- Evans, G. (1977). Pronouns, quantifiers, and relative clauses (I). *Canadian Journal of Philosophy* 7(3), 467–536.
- Fiengo, R. and May, R. (1994). *Indices and Identity*, vol. 24. Cambridge, MA: MIT Press.
- Fine, K. (2003). The role of variables. *The Journal of Philosophy* 100(12), 605–31.
- Fine, K. (2007). *Semantic Relationism*. Oxford: Blackwell Publishing.
- Frege, G., Ebert, P., Rossberg, M., and Wright, C. (1893/2013). *Gottlob Frege: Basic Laws of Arithmetic*. Oxford: Oxford University Press.
- Gauker, C. (2014). How many bare demonstratives are there in English? *Linguistics and Philosophy* 37(4), 291–314.
- Georgi, G. (2015). Logic for languages containing referentially promiscuous expressions. *Journal of Philosophical Logic* 44(4), 429–451.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (eds.), *Syntax and Semantics: Speech Acts*, vol. 3. New York: Academic Press, pp. 41–58.
- Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy* 14(1), 39–100.
- Harbour, D., Adger, D. and Béjar, S. (2008). *Phi Theory: Phi-Features Across Modules and Interfaces*, vol. 16. Oxford: Oxford University Press.
- Haug, D. T. T. (2014). Partial dynamic semantics for anaphora: Compositionality without syntactic coindexation. *Journal of Semantics* 31(4), 457–511.
- Heim, I. (1982). The semantics of definite and indefinite noun phrases, PhD thesis, University of Massachusetts Amherst.
- Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*, Oxford: Blackwell Publishers.
- Higginbotham, J. (1980). Pronouns and bound variables. *Linguistic Inquiry* 11(4), 679–708.
- Higginbotham, J. (1983). Logical form, binding, and nominals. *Linguistic Inquiry* 14(3), 395–420.

- Hunter, J. (2013). Presuppositional indexicals. *Journal of Semantics* 30(3), 381–421.
- Hunter, J. (2014). Structured contexts and anaphoric dependencies. *Philosophical Studies* 168(1), 35–58.
- Jacobson, P. (1999). Towards a variable-free semantics. *Linguistics and Philosophy* 22(2), 117–185.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum, University of Amsterdam.
- Kaplan, D. (1986). Opacity. In L. Hahn and P. Schilpp (eds.), *The Philosophy of W.V. Quine*. Oxford: Oxford University Press, pp. 229–88.
- Kaplan, D. (1989a). Demonstratives. In J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, pp. 481–563.
- Kaplan, D. (1989b). Afterthoughts. In J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, pp. 565–614.
- Lewis, D. (1970). General semantics. *Synthese* 22(1), 18–67.
- Montague, R. (1968). Pragmatics. In R. Klibansky (ed.), *Contemporary Philosophy: A Survey*, Vol. 1, Florence. La Nuova Italia Editrice, pp. 102–22. (Reprinted in *Formal Philosophy: Selected Papers of Richard Montague*, 1974).
- Pickel, B. and Rabern, B. (2016). The antinomy of the variable: A Tarskian resolution. *Journal of Philosophy* 113(3), 137–70.
- Quine, W. (1940/1981). *Mathematical Logic*, revised edn. Cambridge, MA: Harvard University Press.
- Rabern, B. (2013). Monsters in Kaplan's logic of demonstratives. *Philosophical Studies* 164(2), 393–404.
- Rabern, B. and Ball, D. (forthcoming). Monsters and the theoretical role of context. *Philosophy and Phenomenological Research*.
- Radford, A. (2004). *Minimalist Syntax: Exploring the Structure of English*. Cambridge: Cambridge University Press.
- Radulescu, A. (2012). The logic of indexicals, PhD thesis, UCLA.
- Radulescu, A. (2015). The logic of indexicals, *Synthese* 192(6), 1839–60.
- Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford: Clarendon Press.
- Rooryck, J. and vanden Wyngaerd, G. J. (2011). *Dissolving Binding Theory*. Oxford: Oxford University Press.
- Russell, B. (1905). On denoting. *Mind* 14(56), 479–93.
- Safir, K. J. (2004). *The Syntax of Anaphora*, Oxford: Oxford University Press.
- Salmon, N. (2002). Demonstrating and necessity. *The Philosophical Review* 111(4), 497–537.
- Scott, D. (1970). Advice on modal logic. In K. Lambert (ed.), *Philosophical Problems in Logic: Some Recent Developments*, Dordrecht: D. Reidel, pp. 143–73.
- Stojnic, U., Stone, M., and Lepore, E. (2013). Deixis (even without pointing). *Philosophical Perspectives* 27(1), 502–25.
- Tarski, A. (1936). The concept of truth in formalized languages. *Studia Philosophica* 1(4), 261–405.
- Vermeulen, C. (2000). Variables as stacks. *Journal of Logic, Language and Information* 9(2), 143–67.
- Yalcin, S. (2012). Introductory notes on dynamic semantics. In D. Fara and G. Russell (eds.), *Routledge Companion to the Philosophy of Language*. London: Routledge, pp. 253–79.