Same, Models and Representation

Peter Lasersohn
University of Illinois at Urbana-Champaign

1. Introduction

What is the relation between *models*, as used in model-theoretic semantics, and the "world" which models represent? More specifically, let us consider the question of whether a single individual, event, time or other element in a model might be used to represent more than one individual, event, time or other object in the real world.

It should be understood that the question here is not whether groups of objects may be represented with some sort of "plural individual" in the style of Link (1983); rather, the idea is to use a model in which pragmatically irrelevant distinctions between individuals are ignored, so that the model "conflates" more than one real-world entity into a single model-theoretic entity, representing them as though they were the same object. This idea was argued for most explicitly and systematically, to my knowledge, in Nunberg (1984); though related ideas have been hinted at or suggested briefly or casually in a number of other places.

The present paper provides arguments against this technique, as developed by Nunberg, and suggests an alternative approach to dealing with cases where the distinction between real-world individuals is pragmatically ignorable, based on the device of "Pragmatic Halos" presented in Lasersohn (1999).

In considering this issue, we will also have to consider exactly what the point is of using models in model-theoretic semantics; and although this seems like a very elementary issue, and one which lies at the foundation of our whole technique, I think it is still an issue about which there is a great deal of obscurity and differences in thinking. To a surprising degree, semanticists often use essentially identical formalism, yet understand that formalism very differently.

2. The Same – Yet Somehow Different

Nunberg points out that there is an interesting flexibility in the interpretation of adjectives such as *same* and *different*. For example, Sentence (1) can mean either that Enzo now drives the very same car token as I used to drive, or that he drives a car of the same type as I used to drive – presumably meaning in this case, a car of the same model, or the same year and model:

(1) Enzo drives the same car I used to drive.

A similar effect may be observed in Example (2), which can mean either that Enzo and I drive the very same car token, or just that we both drive Ford Falcons:

Brendan Jackson and Tanya Matthews (eds), *SALT X* 83-97, Ithaca, NY: Cornell University.

(2) I drive a Ford Falcon. Enzo drives the same car.

Looking just at examples like these, the most natural analysis would probably be to claim that noun phrases like *the same car* are ambiguous, between a reading requiring token-identity, and a reading merely requiring type-identity, and paraphrasable as something like "a car of the same type."

But as Nunberg points out, an analysis which appeals to this kind of ambiguity faces a number of difficulties. First, and most interestingly for our purposes, it seems to leave as a mystery why it is that you cannot say something like (3):

(3) ?A Ford Falcon was heading south on U.S. 101, went out of control, and crashed into the same car.

It would make perfect sense to say that a Ford Falcon went out of control and crashed into a car of the same type, but for some reason you cannot use the phrase *the same car* here to mean "a car of the same type." It's a very interesting fact.

Second, we find examples like (4), which I have altered slightly from Nunberg's original example:

(4) Otto has been carrying around the same book as he voted to ban last year.

If this use of *the same* depends on a type-token ambiguity in the noun phrase, it should make sense to ask which reading we have in this sentence. But of course what Otto voted to ban was a type, and what he has been carrying around is a token, so it doesn't seem that we can't make a coherent choice here without saying something extra.

Neither of these, perhaps, is really a knock-down argument, and it is easy to imagine analyses of (4), at least, that explain it by positing hidden operators that shift between types and tokens. It is less clear that such an approach can explain the oddity of (3), however; and in any case Nunberg offers a different kind of analysis, which seems to me, at least, to be more interesting and more explanatory than just positing hidden operators – or at least it would be if it worked.

3. Constructing Individuals on a Pragmatically Limited Property Set

Nunberg's analysis is based on the idea that assertions of sameness or difference are always made relative to some particular conversational purpose, and that distinct objects may sometimes be equivalent to one another in their contribution to that conversational purpose. For example, in comparing the body styles of different cars, one Ford Falcon is as good as another, since they all have identically styled bodies; so in a conversation on this topic we might count all Ford Falcons as equivalent to one another and hence refer to any of them as the "same car" as the others.

In contrast, if we are trying to determine who has responsibility for an accident in which one Ford Falcon crashes into another, this is a conversational purpose for

which identical body styling does not produce pragmatic equivalence, so we would not refer to two Ford Falcons as the "same car" in such a conversation.

Nunberg suggests that what makes two objects equivalent for a particular conversational purpose is that they differ only with respect to properties which are irrelevant to that purpose. Ford Falcons differ from one another in all sorts of ways: color, location, age, etc.; but it is easy to imagine conversations in which all these properties are irrelevant to the issues at hand.

Therefore, Nunberg suggests, in giving a model-theoretic semantics for such discourses, we might employ models in which these pragmatically irrelevant properties are simply not represented.

And this is not too surprising an idea. It is commonplace to assume that quantificational sentences, for example, are interpreted relative to a domain of quantification from which pragmatically irrelevant individuals are excluded; and of course each model has a domain of quantification as one of its main components, so we may regard the intended model of a given discourse as not representing pragmatically irrelevant individuals.

Likewise, we might suggest that pragmatically irrelevant times, or events, or whatever, are not represented – and in this light, a suggestion that pragmatically irrelevant properties not be represented seems very natural.

Now, Nunberg suggests, we can "eliminate the universe" in the manner of Keenan (1982) – that is, we don't represent individuals as primitive elements of our models at all, but define them in terms of properties. The technique is well-known: Intuitively, we just replace each individual with its extensional property set. (Of course technically, we don't start with the individuals and then replace them; we just start with the properties, assume boolean operations on them, and then define an individual as a proper principle ultrafilter in the resulting algebra.)

The upshot of this is that any number of distinct individuals in the intuitive sense might correspond to a single individual in the technical sense. In particular, if two individuals differ from one another only in pragmatically irrelevant properties, those properties don't get represented in the relevant model and aren't used in the construction of "individuals" in that model; the real-world individuals might have precisely the same set of pragmatically relevant properties, and therefore correspond to the same ultrafilter – that is, the same individual, in our technical sense – in the model-theoretic representation.

Now, Nunberg suggests, we just define the words *same* and *different* as requiring sameness and difference at the level of model-theoretic representation – in effect allowing us to refer to distinct real-world individuals as "the same *X*" in cases where they differ only in pragmatically irrelevant properties, but prohibiting us from doing so in cases where the individuals differ in some respect which is relevant to our conversational purpose.

It is now possible to see why (3) is anomalous. Nunberg suggests that the kind of discourse where one might expect a sentence like this to occur – say, a police accident report – would normally be a discourse in which the distinction between the two cars is crucial to the conversational purpose – which might be, for example, determining responsibility for the accident. That is, the cars differ in some

pragmatically relevant property, so they would be represented by different property sets in the model, so the use of *same* is not licensed.

Nunberg points out some interesting additional patterns in number morphology that might find an explanation in an analysis along these lines: For example, consider Example (5):

(5) Otto saw an interesting book in Dalton's window yesterday; when he got home, he discovered that Arabella was giving the same {book / ?books} as Christmas presents.

Here, even though Arabella is giving out multiple copies of the book, the more natural phrasing is with the singular noun. Why? On Nunberg's analysis, we can attribute this to the fact that the appearance of the word *same* in this example indicates that all those books are represented model-theoretically as a single individual. If instead we just interpreted *same* as meaning *of the same type*, there is no reason why we should expect that *the same books* in this example would be any more anomalous than *books of the same type*, as in (6), which is perfectly acceptable:

(6) ...he discovered that Arabella was giving books of the same type as Christmas presents.

Nunberg even seems to imply that the same kind of analysis should explain the definiteness of the phrase *the postman's leg* in Example (7):

(7) My dog bit the postman's leg.

Of course the postman probably has more than one leg, so why the definite here? The idea seems to be that the distinction between the two legs is not pragmatically relevant, so they can get represented by a single individual in the model – and it is at this level that the uniqueness presuppositions for definites gets satisfied.

It's worth noting in this connection that noun phrases with *same* are required to be definite, as shown in (8).

(8) a. The same X
b. *A same X

4. Models as a Level of Representation

Nunberg's analysis is challenging, and interesting from a broader theoretical perspective, because it seems to turn models into a significant level of representation, whose relation to the world they represent is not particularly direct or straightforward. In this view, the "individuals" in a model, even in the relevant, intended model for a particular discourse, are not identified with the real-world individuals we talk about, nor do they even stand in one-one correspondence with those individuals. Instead,

a substantive issue arises about which individuals in the model correspond with which individuals in the world, and this correspondence relation becomes something we need to study and theorize about, rather than something trivial and automatic.

In fact Nunberg goes further, and suggests that if we adopt this view of models, the reconstruction of individuals as property sets becomes superfluous – once we regard individuals in a model as representations of real-world individuals rather than as the real-world individuals themselves, we can return to the idea of treating them as formally primitive if we want, so long as they are differentiated only by pragmatically relevant properties.

Nunberg's analysis is, perhaps, an extreme case, but analogous ideas come up in a less dramatic way with some frequency. For example, Link (1987) outlines a technique for dealing with issues of "granularity" in the representation of events. Issues of this kind come up, for example, in defining aspectual classes.

Take the case of Mozart's death, described in (9):

(9) Mozart died on the 5th of December, 1791.

In most respects, *die* behaves like an achievement verb, describing an atomic event with no temporal parts. But as Link puts it, "There is ... a level of description on which Mozart's death is composed of a more or less complex series of events, e.g., the physical processes involved in his dying."

Link therefore suggests that we employ a whole series of different event lattices, each one displaying the part/whole structure of events at a different granularity of representation. Homomorphisms map the more coarse-grained structures into the more fine-grained structures, so that whatever structure exists at the coarser levels is preserved in the finer levels. But it is entirely possible that an element might be atomic at a coarse level of representation, even while its image under one of these homomorphisms might have proper parts at the finer levels of representation.

Although Link does not formalize this system in much detail, the idea seems to be that these event lattices should be components of our models, in the usual sense from model-theoretic semantics. It seems, then, we must regard the event lattices in our models as representations, whose relation to the real events and part/whole structure they represent may not be particularly direct or straightforward.

A related perspective is laid out in some detail in Zimmerman (1999). Zimmerman writes (p. 540):

> *"...the individuals* [*in models*] *are objects of pure set theory and do not lead lives in any literal sense. Similarly, the situations are abstract objects of set theory and none of them coincides with an actual situation. Of course, these abstract objects can and should be thought to* represent *individuals and situations, but it is important to realize that they are neither."*

Once we regard the individuals and situations in a model – even the "intended" model of a particular discourse – as representations, the issue naturally arises of how to relate them to the real individuals and situations they represent, and Zimmerman discusses this in some detail. For entities with names, we let the real bearer of the name correspond to the object assigned as the denotation of the name in the model; and for objects without proper names, we set up the correspondence by relating them to objects with names, as in (10), which relates people's noses to their model-theoretic representations. (Here M_0 is one of the intended models for English, with sets of individuals D_0 and situations S_0, and interpretation function F_0; '~' represents the correspondence relation between model-theoretic objects and the real-world things they represent.)

(10) For any individuals $\boldsymbol{x}$, situations $\boldsymbol{s}$ and M_0-entities $x \in D_0$ and $s \in S_0$:
if $x \sim \boldsymbol{x}$ and $s \sim \boldsymbol{s}$, then $F_0(\textbf{nose}')(s)(x) \sim \boldsymbol{x}$'s nose in $\boldsymbol{s}$.

Once we recognize the need for correspondence rules of this sort, the argument goes, meaning postulates become otiose, and the role of models in semantic theory becomes very limited.

There is nothing here about model-theoretic entities corresponding to multiple real-world entities, but the fundamental perspective of model-theoretic entities as representations is a crucial part of the argument.

But what reason do we have for thinking of models in this way? I share many of Zimmerman's concerns about models and meaning postulates, but I think this question of correspondence between model-theoretic entities and their real-world counterparts is a red herring.

I suspect maybe some people would justify the representational view of models this simply: Models are ordered pairs, or *n*-tuples of some kind; the real world is not an ordered pair or *n*-tuple; therefore models are at best only a representation of the world, not the world itself. And once we start thinking of models as representations, it seems almost automatically to open these issues concerning the representation relation that lead to concerns of the type that Zimmerman expresses.

However, there is a big difference between saying that models are representations, and saying that the elements of which models are composed are representations.

If we define a model as an ordered pair containing a non-empty set and a function meeting certain conditions, then *any* non-empty set and *any* function meeting those conditions will qualify. I, for one, have no problem with the idea of real, concrete, physical objects being members of sets, and so some of our models will have real, concrete, physical objects as members of their domains, and not just "abstract objects of pure set theory." If we can accept this, then there is no obstacle to the idea of a model whose interpretation function assigns basic expressions the things they really denote, and not just abstract representations of those things. In fact I would say that there will be such models, and that the whole question of how to fix a representation relation between model-theoretic entities and their real-world counterparts is a spurious issue.

None of this says anything against the idea that there is a level of representation which intervenes between expressions in a language and their denotations in the world – something like Discourse Representation Structures, for example. But I think it is a serious mistake to think of model-theoretic denotations as playing that role. There really is a major difference here, even though people sometimes talk about Discourse Representation Structures as "partial models." There is some justification to this kind of terminology, but Discourse Representation Structures don't play at all the same role in semantic theory that models classically play – that is, they don't play the role of M in definitions like (11), for logical consequence:

(11) φ is a logical consequence of ψ iff for all M: if ψ is true in M then φ is true in M.

I assume this is what we mean by a model – it's something that plays the role of M in this sort of rule; and unless we make use of this sort of rule, explicitly or implicitly, there is no point in using models in the first place.

If we were just defining truth, for example – ordinary truth, not logical truth – we would have no motivation for considering the denotations of expressions across a range of models. We might still relativize denotations to times, worlds, or other indices, but there would be no reason to bring models into the picture at all – we would just assign truth values based on the denotations which expressions actually have relative to those times, worlds, or other indices. This is the technique in Lewis (1972), for example, and many other places.

In fact, even if we are concerned with entailment relations, and not just truth, we can define some kinds of entailment modally, in terms of the inclusion relation on sets of possible worlds, again without bringing models into the picture. We really only need to appeal to models if we are concerned with defining a logical consequence relation which differs in some relevant respect from the modal necessitation relation – for example if we think of logical consequence as licensing arguments based specifically on their syntactic form.[1]

But if there is no motivation for using models when our main concern is just with truth, or with modally-definable entailment, there certainly can't be any motivation in such cases for exploiting a putative distinction between an expression's denotation in a model, and the real-world entity or entities that that model-theoretic denotation represents.

Of course, whatever ontological assumptions we need to make in the assignment of ordinary truth values will have an effect on how we construct models, once we turn to the definition of logical consequence and related notions; I'm not trying to argue against *that*. But the reverse kind of case – where the denotation of an expression relative to a model is different from the denotation we would assign if we were not using models – that kind of case, it seems to me, can only be motivated on the grounds that it is needed for the definition of logical consequence, logical equivalence, logical truth, and so on, and not in the analysis of ordinary truth and falsity.

It is interesting to note in this respect that Nunberg's original analysis of *same* and *different* was entirely motivated by problems in the ordinary assignment of truth values, not in the definition of logical consequence or other logical notions – in fact, logical consequence is hardly mentioned in the whole article. So methodologically, I don't think there is any reason to appeal to models in developing a solution – certainly no reason to appeal to distinction between model-theoretic denotations and their real-world correspondents; and any solution which crucially makes such an appeal is on shaky ground.

5. An Effability Problem

Of course, the argument just given is made on a purely conceptual, methodological basis, and may not be convincing to semanticists who don't already share the same methodological assumptions. After all, part of Nunberg's point seems to be methodological, and if we regard his paper primarily as an argument for a particular methodological approach in constructing models, we can hardly argue against it simply by assuming the kind of methodology he argues against.

So let us turn now to a different problem, of a more technical nature, which I think will pose trouble for Nunberg's analysis even if we concede this use of models as appropriate.

To review: the basic idea is that the domain of properties in a model is pragmatically limited in the same way as the domain of individuals usually is in more conventional models, so that only pragmatically relevant properties are represented. Then we use this limited set of properties to individuate our other entities: collapsing together all those individuals which share all the same properties (from among those represented in the model), and keeping separate those individuals which differ in at least one of these properties.

Of course the denotations of predicates in a given model are also going to come from the set of properties represented in the model. And this is what is going to cause us problems. Suppose I am the sole owner of Ford Falcon Number One, and Enzo is the sole owner of Ford Falcon Number Two. Then Ford Falcon Number One has at least one property which is not shared with Ford Falcon Number Two, namely λx[I own x]; and of course Ford Falcon Number Two has a property that is not shared with Ford Falcon Number One, namely λx[Enzo owns x].

Now, suppose these properties are pragmatically relevant to the issue at hand in our discourse. Then these two properties will be represented in the model; the two cars will not correspond to the same property set, so they will be represented separately, and we should not be able to refer to my Ford Falcon and Enzo's Ford Falcon as "the same car." So far so good.

But now suppose that these two properties are pragmatically irrelevant. In this case, they will not be represented in the model. So the two cars will collapse together, and get represented with a single model-theoretic individual, which is what we want. But, since these properties aren't represented in the model, and the denotations for all our expressions have to be drawn from those assigned in the

model, we will have no way of expressing these properties. So we should not be able to say anything like (12), since the italicized portion would have to denote one of these unrepresented properties:

(12) I own a Ford Falcon. The same car *is owned by Enzo*.

But in fact this is a perfectly interpretable, coherent discourse, so it appears that there is something wrong. Nunberg's analysis predicts that if we refer to two objects as the same, we should not be able to express any properties which differentiate between them, at least in the same part of the discourse, and that is too strong.

6. Sameness in a Pragmatic Halos Framework

Despite the problems discussed above, I think there is something fundamentally right about Nunberg's analysis – in particular, I will follow him in saying that *Enzo drives the same car as I used to drive* is acceptable, not because of any type/token ambiguity, but because, in context, we are allowed to pragmatically ignore the distinction between the two cars. The problem is in how we represent individuals between which the distinction is being ignored. We should not represent them in a way which prohibits all mention of properties which distinguish them.

Lasersohn (1999) suggested an analysis for a very different kind of case where the distinction between objects is pragmatically ignorable, using a rather different formalism for representing such cases. Here, I would like to suggest that this formalism be adapted for the sort of examples that Nunberg addressed.

The original concern in Lasersohn (1999) was with examples like (13)a. and b.:

(13) a. Mary arrived at three o'clock.
b. Mary arrived at exactly three o'clock.

We can certainly say something like (13)a. even if Mary arrived, say, five minutes late. But it seems wrong to claim that the sentence is actually *true* in such a situation. On the contrary, the sentence is true only if Mary arrives at three o'clock; it is false if she doesn't arrive till later.

We can use the sentence felicitously if she arrives at 3:05, not because it is true, but because certain minor kinds of falsehood are pragmatically permissible. The distinction between 3:00 and 3:05 may not be relevant to the purposes of our discourse, so we can ignore it, counting the sentence as close enough to true for its context, even if not really true.

This raises an interesting puzzle about the difference in meaning between (13)a. and (13)b.: If (13)a. is true only if Mary arrives right on the button, then it would seem to be truth conditionally equivalent to (13)b. But intuitively, there seems to be a meaning difference here. The question is how to capture it.

I believe that (13)a. and (13)b. are in fact truth conditionally equivalent, and that the difference in meaning is not truth conditional, but in the degree of deviation from the truth which the two sentences permit without producing pragmatic infelicity: (13)a. permits a greater degree than (13)b., so (13)a. can be felicitous even if Mary arrives so much later than 3:00 that (13)b. would be infelicitous.

Formally, we can account for this by associating with each expression, relative to a given pragmatic context, not only its denotation, but also a set of items understood to differ from the denotation only in ways which are pragmatically ignorable in that context. For example, the denotation of a phrase like *three o'clock* will be the actual time, three o'clock, but we also associate with this phrase a set of times which differ from three o'clock only in pragmatically ignorable ways – for example these might be times just a few minutes before or after three o'clock, or short intervals which include three o'clock but also a little time on either side. We can call that set the *pragmatic halo* of the expression. Notationally, we may represent the halo of an expression as shown in (14).

(14) $H_C(\alpha)$ = the 'pragmatic halo' of α in context *C* (i.e. the set of things which differ from $[\![\alpha]\!]_C$ only in ways which are pragmatically ignorable in *C*)

In the default case, we can calculate the halo of a complex expression compositionally just by applying our normal semantic rules to all combinations of elements drawn from the halos of its parts. For example, given some complex expression with parts α and β, where the denotation of the complex expression is fixed by rule as the value returned by applying the function denoted by α to the object denoted by β, the halo of the complex expression will be the set of all values returned by applying the various functions in the halo of α to the various objects in the halo of β, as shown in (15).

(15) Suppose $H_C(\alpha) = \{f, g, h\}$ and $H_C(\beta) = \{a, b, c\}$, where by rule: $[\![\alpha\beta]\!]_C = [\![\alpha]\!]_C([\![\beta]\!]_C)$. Then $H_C(\alpha\beta) = \{f(a), f(b), f(c), g(a), g(b), g(c), h(a), h(b), h(c)\}$

This example assumes functional application, but the basic principle can be straightforwardly adapted to other sorts of semantic operations as well; see Lasersohn (1999) for details.

This procedure gives the result that a sentence might have the value 'true' in its halo even if it has 'false' as its denotation – this would be the case for (13)a., for example, if Mary did not arrive at three o'clock, but only at one of the times in the halo of three o'clock, say, five minutes after three. In such a case, we would count the sentence as "coming close enough to true for its context," as indicated in (16):

(16) φ comes close enough to true enough for *C* iff $1 \in H_C(\varphi)$.

Now we can explain the function of the word *exactly* in (13)b. in terms of halos. It has no truth conditional effect, so that the phrase *exactly three o'clock* has

exactly the same denotation as *three o'clock*. What *exactly* does is contract the halo, so that the halo of *exactly three o'clock* will just contain the actual time three o'clock, plus maybe some other very close times, but will not include the more outlying elements of the halo of *three o'clock* – for example, maybe not 3:05. As a result (13)b. may not come close enough to true for its context even in cases where (13)a. does.

Like Nunberg's technique, pragmatic halos give us a way to represent examples where the distinction between objects is pragmatically ignorable. Unlike Nunberg's technique, it is relatively silent on the reasons for this ignorability – if we want, we can retain Nunberg's idea that only certain properties may be relevant in a given context, and that the distinction between individuals that differ only in irrelevant properties is ignored – though this is not required by the formalism.

But even if we retain this idea, it does not commit us to using just a single model-theoretic entity to represent multiple individuals; nor does it commit us to the claim that any property which distinguishes between individuals is completely inexpressible in a discourse which treats those individuals as equivalent.

For example, Enzo's car might be in the halo of my car – but this just means that it is included in a set which we associate with the phrase *my car* – not that we represent the two cars with a single individual in the model. And since the two cars need not be represented with a single individual, they don't have to correspond to precisely the same property set, so we don't have to say that properties like being owned by me or being owned by Enzo are simply not represented in the model and therefore can't serve as the denotations of linguistic expressions.

Although this technical problem is eliminated, we do still have the more general conceptual or pragmatic problem of why, if these properties are pragmatically irrelevant, anyone would bother to mention them, even if they are formally available to serve as denotations. I have little to say in this regard, except to suggest that it is probably too simplistic to think that we can get away with a simple two-way distinction between relevant and irrelevant properties, corresponding to representation or lack of representation in our models. Instead, relevance is always relative to a particular conversational purpose, and speakers might be pursuing multiple conversational purposes at the same time, even in a single sentence, with each purpose giving rise to its own sets of relevant and irrelevant properties. Details of ownership might be irrelevant in a particular discourse for the purposes of determining whether my car and Enzo's car count as "the same car," yet still be relevant enough to mention for some other purpose, allowing us to say things like *I own the same car as Enzo*. The crucial thing is that even in mentioning ownership, there is no incoherency in speaking as though the two cars were identical, whereas we do get an incoherent result if we speak as though two cars were identical when describing one as crashing into the other, as in (3), since – obviously – a car cannot crash into itself.

It is interesting to note in this connection that the incoherency effect really persists only through the clause; after that it becomes possible to refer to the cars in a way which absolutely requires them to be distinct, even if they were previously described as the same, as in (17):

(17) I drive a Ford Falcon, and Enzo drives the same car. His was heading south on U.S. 101, went out of control, and crashed into mine.

I don't have a detailed explanation for this fact, but suspect that it is a matter of presupposition accommodation. The second sentence presupposes there are two separate cars; we can accommodate that presupposition, in effect creating a new pragmatic context, with its own halo assignments, where the cars are not in each other's halos; and we interpret the second sentence relative to this new context. We cannot do the same thing in (3), however, because the clause which describes the crash is the same one that describes the cars as the same, and the word *same* requires one car to be in the halo of the other.

Turning now more specifically to the semantics of *same* in this kind of analysis, I will concentrate here just on problems of the sort Nunberg addressed, setting aside the "quantifier dependent" reading of *same* that Carlson (1987), Moltmann (1992), Beck (2000) and others have discussed.

We have a couple of options in formulating an analysis. Probably the simplest one – though probably not the most intuitive one – would be to assume that *same* is indexed to an antecedent phrase, and then define it as in (18):

(18) $\mathbf{same}_i' = \lambda P \lambda y[P(y) \,\&\, y = x_i]$

In fact I think this is the semantics most of us would come up with if we didn't think about examples like Nunberg's; it requires strict identity, at least as far as truth conditions are concerned. So (19)a. would get the semantics indicated in (19)b., assuming for simplicity that the anaphora is handled by existential closure.

(19) a. I drive a Ford Falcon$_1$. Enzo drives the same$_1$ car.
b. $\exists x_1[\mathbf{ford\text{-}falcon'}(x_1) \,\&\, \mathbf{drive'}(\mathbf{I'}, x_1) \,\&\, \mathbf{drive'}(\mathbf{enzo'}, \iota y[\mathbf{car'}(y) \,\&\, y = x_1])]$

Of course this requires that Enzo drive the very same car token that I do, which is not the reading we're trying to account for. So if we take this option we have to claim that in the situation where Enzo drives a physically distinct, but equivalent car to mine, the sentence is literally false. Of course if we make use of the device of pragmatic halos, it might still come close enough to true for its context – and we get this effect straightforwardly just by assuming that Enzo's car is in the halo of my car – but in effect we'd be saying that the relevant reading is produced via a kind of pragmatic reinterpretation, not as part of the truth conditions. In fact, we would be treating this example in completely parallel fashion to *Mary arrived at three o'clock* – as something that is literally false, but in a minor, pragmatically tolerable way.

This kind of approach might appeal to proponents of Radical Pragmatics, and personally, I'm not so sure it's wrong. But it does force a fairly large separation between our naive intuitions and the reading which the truth conditions give us, so I think it is worth thinking through whether there might be an alternative approach, that people with narrower throats than mine could still swallow.[2]

If the goal is to treat discourses like (19)a. as literally true even if Enzo and I merely drive distinct cars of the same type, we can do this by treating sentences containing *same* as involving quantification over the members of the halo of the antecedent. That is, even at the level of truth conditions, we have quantification over the halo – so in effect, *same* converts pragmatic "slack" into semantic content.

As a first attempt at this, we might define *same* as in (20):

(20) $\mathbf{same_i'} = \lambda P \lambda y[P(y) \,\&\, y \in \mathbf{H}(x_i)]$ (where for all contexts C, $[\mathbf{H}]_C = H_C$)

In effect, this fixes the denotation of *same car* as the intersection of the set of cars with the halo of the antecedent. This would be fine, except that the halo might contain any number of cars, so this set might easily turn out not to be singleton. But recall the pattern in (8): noun phrases with *same* are obligatorily definite. This is similar to the pattern in (21); noun phrases with *only* are obligatorily definite (except of course in the fixed phrase *an only child*):

(21) a. the only answer
b. *an only answer

I believe the reason for the definiteness is the same in both cases: *same*, like *only*, should never give us more than a singleton set when it combines with a singular noun. Unfortunately, (20) doesn't do that.

To explain the definiteness here, I will borrow a technique from the approach to *in*definites presented in Reinhart (1997) and Winter (1997), and suggest that rather than simply taking the intersection of the denotation of the head noun and the halo of the antecedent, *same* involves the selection of a particular member of the halo. Formally, we represent *same* using a variable over choice functions, as in (22):

(22) $\mathbf{same_i'} = \lambda P \lambda y[\mathrm{CH}(f) \,\&\, P(y) \,\&\, y = f(\mathbf{H}(x_i))]$
(where $\mathrm{CH}(f)$ iff $\forall X \in dom(f)$: $f(X) \in X$)

Here f picks out a particular member of the halo of the antecedent, and – provided that thing is a car – the phrase *same car* will denote the singleton set containing it.

However, f should be understood as a free variable over choice functions, not as a constant picking out a particular choice function. As a free variable, it gets bound by existential closure, with the result that the discourse in (19)a. gets represented as in (23):

(23) $\exists x_1, f[\mathbf{ford\text{-}falcon'}(x_1) \,\&\, \mathbf{drive'}(\mathbf{I'}, x_1) \,\&\, \mathrm{drive'}(\mathbf{enzo'}, \iota y[\mathrm{CH}(f) \,\&\, \mathbf{car'}(y) \,\&\, y = f(\mathbf{H}(x_1))])]$

Since different values for f will pick out different cars from the halo of x_1, we get the effect of existential quantification over the halo, even while guaranteeing that the common noun argument to the determiner *the* will never denote more than a singleton set.

Admittedly, this is a somewhat artificial means of assuring definiteness, and it is perhaps worth noting that our initial definition back in (18) gave this effect automatically, without any appeal to choice functions.

7. Conclusion

Same expresses identity between an individual in the denotation of its common noun complement, and an individual selected from the halo of its antecedent. Since these individuals are not conflated in model-theoretic representation, properties which distinguish them may still be expressed, and models retain their traditional status and use.

Endnotes

*Thanks to Chris Barker and to the audience at SALT for useful comments. Errors are my own.

[1] From this perspective, an argument like (i) is logically valid because it is of the general form in (ii):

(i) Every dog is a mammal;
Every mammal is an animal;
Therefore, every dog is an animal.

(ii) Every X is a Y;
Every Y is a Z;
Therefore, every X is a Z.

The form in (ii) is valid because every assignment of values to X, Y and Z which makes the premises true also makes the conclusion true. But why pass through the intermediate stage of replacing the nouns in (i) with variables? We might as well just consider all possible assignments of values to the nouns in (i) directly. In essence, this is what I think we are doing in defining logical consequence in terms of models; it is a quite different enterprise from checking whether every possible world in which the premises are true is also a world in which the conclusion is true.

For arguments that logical truth is distinct from the necessary truth, see e.g. Kaplan (1989), Zalta (1988).

[2] Intuitions vary somewhat from example to example. Consider the case of my old friend David: He dated a woman; they broke up; when I saw him again a year later he was dating a different woman, who resembled the first one to an astonishing degree. If this pattern were repeated several times, we might very well say *David just keeps dating the same woman over and over*; but the intuition is that the sentence is literally false (albeit felicitous).

References

Beck, Sigrid (2000) 'The Semantics of *Different*: Comparison Operator and Relational Adjective', *Linguistics and Philosophy* 23.2.101-139.

Carlson, Greg (1987) 'Same and Different: Some Consequences for Syntax and Semantics', *Linguistics and Philosophy* 10.4.531-565.

Kaplan, David (1989) 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals', J. Almog, et al., eds., *Themes from Kaplan*, Oxford University Press, Oxford, pp. 481-563.

Keenan, Edward (1982) 'Eliminating the Universe', in *Proceedings of the First West Coast Conference on Formal Linguistics*, D. Flickinger, et al., eds. Dept. of Linguistics,. Stanford University, pp. 71-82.

Lasersohn, Peter (1999) 'Pragmatic Halos', *Language* 75.3.522-551.

Lewis, David (1972) 'General Semantics', in *Semantics of Natural Language*, D. Davidson and G. Harmon, eds., D. Reidel, Dordrecht, pp. 169-218.

Link, Godehard (1983) 'The Logical Analysis of Plural and Mass Terms: A Lattice-Theoretical Approach', in *Meaning, Use and Interpretation of Language*, Rainer Bäuerle, Christoph Schwarze and Arnim von Stechow, eds. Walter de Gruyter, Berlin, pp. 302-323.

Link, Godehard (1987) 'Algebraic Semantics of Event Structures', in *Proceedings of the Sixth Amsterdam Colloquium*, J. Groenendijk, et al., eds., ITLI, Amsterdam, pp. 243-262. Reprinted (1998) in Godehard Link, *Algebraic Semantics in Language and Philosophy*, CSLI Publications, Stanford, pp. 251-268.

Moltmann, Friederike (1992) 'Reciprocals and *Same/Different*: Towards a Semantic Analysis', *Linguistics and Philosophy* 15.4.411-462.

Nunberg, Geoffrey (1984) 'Individuation in Context', in *Proceedings of the West Coast Conference on Formal Linguistics 3*, M. Cobler, et al., eds. Stanford Linguistics Association, pp. 203-217.

Reinhart, Tanya (1997) 'Quantifier Scope: How Labor is Divided between QR and Choice Functions', *Linguistics and Philosophy* 20.4.335-397.

Winter, Yoad (1997) 'Choice Functions and the Scopal Semantics of Indefinites', *Linguistics and Philosophy* 20.4.399-467.

Zalta, Edward (1988) 'Logical and Analytic Truths that are not Necessary', *Journal of Philosophy* 85.2.57-74.

Zimmerman, Thomas Ede (1999) 'Meaning Postulates and the Model-Theoretic Approach to Natural Language Semantics'. *Linguistics and Philosophy* 22.5.529-561.