CrossMark

A puzzle about demonstratives and semantic competence

Jeff Speaks[1]

Abstract My aim in this paper is to lay out a number of theses which are very widely held in contemporary philosophy of language and linguistics, and to argue that, given some extra theses for which I'll argue, they are inconsistent. Some of this will involve going through some very well-trodden territory—my hope is that presenting this familiar ground in the way that I do will help to make plain the problem that I aim to identify.

1 Compositionality and semantic competence

The central aim of (in one use of that term) the theory of meaning is to give a compositional semantic theory for a natural language. We can think of a theory of this sort as doing two things:

- Assigning meanings to the simple expressions of the language.
- Giving rules for deriving the meanings of arbitrary complex expressions from the meanings of the expressions of which they are composed, and the way in which those expressions are combined.

This description of the aim of semantics raises a question: why think that there is any true theory of this sort for languages like English to be given?

This is a question which has a widely agreed-upon answer—one traceable at least to Frege, who wrote that

---

✉ Jeff Speaks
jspeaks@nd.edu

[1] University of Notre Dame, Notre Dame, IN, USA

Springer

the possibility of our understanding sentences which we have never heard before rests evidently on this, that we can construct the sense of a sentence out of parts that correspond to words.[1]

The standard interpretation of Frege's remark is that he is proposing a certain sort of explanation of the fact that we can understand novel sentences. The explanation, roughly, is that we know something about the words of our language, and the way that they are combined, which enables us to figure out the meanings of complex expressions of the language (including sentences). Once one is this far, it is very natural to suggest that what one knows about the simple expressions of the language is their meaning, and what one knows about the way words can be combined in the language is the way in which the meanings of complex expressions depend upon the meanings of the expressions of which they are composed. But if this explanation is correct, there must be some true compositional semantic theory of English for us to know.

The fact that we have the ability to understand novel sentences is sometimes called the *productivity* of our linguistic competence. There is another, distinct but related, fact about our linguistic competence which is also often used in arguments for the existence of a compositional semantic theory, which is typically called *systematicity*. Suppose that we understand some number of complex expressions which are all of the syntactic form f, and that the we list the simple expressions $e_1 \ldots e_n$ which make up those complex expressions. Then the thesis of systematicity says that we can understand any other complex expression which (1) is of form f and (2) is entirely composed of simple expressions from $e_1 \ldots e_n$.

Examples of the phenomenon to which the general thesis of systematicity points are not hard to come by. It is hard, for example, to imagine someone who can understand 'The student wrote a paper' and 'The boy read a novel' who could not also understand 'The student read a paper', 'The boy wrote a novel' etc. Like productivity, systematicity has been used to argue for the existence of a compositional semantic theory for natural languages. Systematicity would be well-explained by the thesis that our semantic competence is due to our grasp of a compositional semantic theory: for if we in general know the meanings of complex expressions of a certain form by computing them on the basis of our knowledge of the meanings and our knowledge of the meaning of the relevant syntactic form, we should expect that we will be able to similarly compute the meaning of any complex expression of that form which consists of simple expressions we understand. And it seems plausible that it is the best explanation, because it is hard to imagine another explanation of comparable plausibility.

Of course, not everyone thinks that productivity and systematicity are to be explained by our literally having knowledge of a semantic theory. And nothing in what follows will depend on the explanation being given in terms of *knowledge*, rather than some other sort of representation. But it is worth emphasizing that for the standard Fregean argument in favor of the existence of a compositional semantic theory to go through, we do need to be making a genuine explanatory claim to the

---

[1] Frege (1914/1980, 79).

effect that our competence is productive and systematic *because* we in some way represent the contents of a compositional semantic theory for the language. If one took a more instrumentalist view, according to which a compositional semantics simply provides a useful model of our semantic competence without genuinely explaining it, it would be hard to see how one could use the nature of our semantic competence as part of an argument for the existence of a compositional semantic theory for a natural language.[2]

I'd like now to introduce a few bits of terminology to slightly regiment the sketch of the standard picture just given:

- So far I've been talking about our knowledge of the meanings of expressions of our language. For reasons shortly to become clear, it will be useful replace this talk of knowledge of meaning with talk about knowledge of *content*.[3] Since it is plausible that our competence with sentences of our language at least involves knowledge of their truth conditions, we can assume that contents are things which are, or determine, truth conditions. I'll otherwise be leaving the notion of content open between competing the competing views of that notion given by Fregean, Russellian, possible-worlds, and other approaches to semantic theory.
- A *compositional theory of content* will be a theory which assigns contents to the finitely many simple expressions of the language and provides finitely many rules which, for any complex expression of the language, tells us how to compute the content of the complex expression in terms of the contents of its parts.
- A compositional theory of content for a language will be *adequate* just in case, for each complex expression of the language, it entails a true claim which gives the content of that expression.[4]
- A compositional theory of content will be *internalizable* just in case it is a theory grasp of which could plausibly explain our knowledge of the contents of the complex expressions about which that theory delivers a verdict.[5] I will assume that a theory which contained infinitely many lexical entries or infinitely many independent rules for deriving the contents of complex expressions from the contents of their parts would not be graspable by us, and hence would not be internalizable.

This leaves a host of questions unanswered (What is it to grasp a theory of this sort? How can we tell if a theory is adequate?). But it will be enough to get started. In the

---

[2] I should also note that there is controversy about how to understand all of the key terms here—'compositional,' 'productive,' and 'systematic.' But nothing in what follows will hang on any of the fine points about which there is controversy.

[3] The reason is that, when it comes to context-sensitive expressions, 'meaning' is ambiguous between 'standing meaning' and 'meaning relative to a context.' More on this below.

[4] What sort of true claim? The answer is that it depends a bit on one's view of semantics. A Davidsonian will think that it is a T-sentence, a possible worlds theorist might think that it should be the possible worlds truth condition of a sentence, and a structured propositions theorist might think that it should be a pairing of sentences and structured propositions. The somewhat awkward wording of the adequacy condition is meant to be neutral on this point.

[5] Of course, the scope of our semantic competence includes, because of our cognitive limitations, only some of the infinitely many constructions about which a compositional semantics entails theorems. I ignore this for simplicity in what follows.

terms just given, I think that the standard picture of the linguistic competence of English speakers can be summarized as follows:

[S&P]   Our knowledge of the contents of complex expressions of English is systematic and productive.

[C]   If our knowledge of the contents of complex expressions of English is productive and systematic, then there is a compositional theory of content for English which is both adequate and internalizable.

2 Compositionality and Kaplan's account of context-sensitivity

The picture so far is plausible enough—but is also, in at least one obvious sense, incomplete. After all, there are some expressions of English which seem to have different contents depending on the occasion on which they are used. The most obvious examples are 'I,' 'here,' and 'now', 'this' and 'that'—but there are others. I'll refer to any expression which has different contents in different contexts as *indexical*, or *context-sensitive*.

Why assume that 'I' has different contents in different contexts? This is obvious if we take the content of 'I' in a context to be the object for which it stands in that context, since it obviously picks out different individuals in different contexts. But, noticing that 'I' always refers to the speaker of the context, one might try to get around this problem by treating 'I' as synonymous with 'the speaker of the context.' But this won't work. As Kaplan pointed out, this sort of view of 'I' would make it a bit hard to understand sentences like 'I wish I were not speaking' and 'In 100 years, I will be dead'—since these would, on this interpretation, respectively express the incoherent wish that the speaker of the context not speak, and assert the bold claim that in 100 years time, the dead will speak.[6]

So let's assume, as is standard, that indexicals like 'I' really do have different contents with respect to different contexts. I now want to point out a parallel between our knowledge of the content of 'I' with respect to different contexts and our knowledge of the contents of complex expressions of English. The parallel might be laid out like this:

Element 1	Element 2	Compound
Simple expressions	Syntactic forms	Complex expressions
Simple indexicals	Contexts	Indexical/context pairs

---

[6] One could get around this problem by letting 'I' mean not 'the speaker of the context,' but rather 'the speaker of this context.' This would lead to other counterintuitive assignments of truth conditions to sentences—but more importantly for our purposes, since 'this' is itself an indexical which presumably has different contents with respect to different contents, it would not provide a way of avoiding the claim that 'I' has different contexts relative to different contexts.

In both cases, we have certain things which have contents which can be thought of as built up out of simpler things—complex expressions built up out of some simple expressions + a mode of combination, and indexical/context pairs built up out of a simple indexical + a context.

The arguments of the last section from productivity and systematicity can be thought of as beginning by noticing a fact about our knowledge of the content of the compound, and inferring from it a claim about our knowledge of certain properties of the elements of which the compound is composed. What I now want to ask is whether these arguments generalize from our knowledge of the contents of complex expressions to our knowledge of the contents of indexicals in contexts (indexical/context pairs).

The argument from productivity does seem to generalize. For just as we somehow manage to figure out the contents of novel complex expressions, we also manage to figure out the contents of novel indexical/context pairs. When I meet someone new, I have no trouble at all figuring out the content of their uses of 'I,' despite the fact that this indexical/context pair is, for me, a novel one. This seems exactly parallel to the fact that I have no trouble at all figuring out the meanings of sentences which I've never before encountered.

Nor is 'I' at all special in this respect. To focus on the example which will interest us later, I have encountered the simple demonstratives 'this' and 'that' as used in any number of contexts. And yet the fact that I have never before encountered 'this' as used in some particular context is no bar to me figuring out what the content of 'this' is in that context. This again seems parallel to my ability to figure out the contents of novel sentences.

Above we explained this fact about our knowledge of the contents of complex expressions in terms of our knowledge of some properties of the elements from which complex expressions are constructed—simple expressions and syntactic forms—which together enable us to derive the meaning of the complex expression. Those properties were, roughly, the contents of the simple expressions, and rules which tell us how to compute the contents of complex expressions of certain forms, given an assignment of contents to the simple expressions of which they are composed. If this reasoning is persuasive in that case, it is hard to see why it should not be so in the present case as well. So it looks like the productivity of our knowledge of the contents of indexicals in contexts should be explained in terms of our knowledge of some properties of simple indexicals and contexts which together enable us to derive the content of that indexical in that context.

The argument from systematicity also seems to generalize. If I can determine the content of one indexical with respect to one context, and another indexical with respect to another context, I can also determine the content of the first indexical with respect to the second context, and vice versa. Hence if the systematicity of our competence with complex expressions is best explained by our grasp of a compositional theory of content, it seems that we should require some parallel explanation of the systematicity of our knowledge of the contents of indexicals in contents.

Fortunately, it is very plausible that the needed explanation of the productivity and systematicity of our knowledge of the contents of indexicals in contexts is

already on hand, thanks to the pioneering work of Kaplan (1977/1989). Let's say, following Kaplan, that *characters* are functions from contexts to contents. Then, one might think, we can explain our knowledge of the contents of indexicals in contexts in terms of our knowledge of the characters of expressions in much the same way as we explained our knowledge of the contents of complex expressions in terms of our knowledge of the contents of the simple expressions which compose them, plus our knowledge of the ways in which those simple expressions are combined. Once we know the character of an indexical, to determine its content with respect to an arbitrary context, all that we need is the information about the context to which that character is sensitive. Which information this is will vary depending on the indexical; in the case of 'I' we will need to know the speaker of the context, and in other cases, other contextual features will be relevant. Though Kaplan does not put things in terms of productivity and systematicity, it is plausible that he wanted characters to play something like this role in explaining our knowledge of the contents of indexicals in contexts—he does, after all, think of them as 'what is known by the competent language user.' (505)

Let's say that a *theory of character* is an assignment of characters to all of the simple expressions of the language. Then (just as above) we can say that a theory of character is *adequate* iff it entails, for any context, a true claim about the content of each indexical in that context, and *internalizable* just in case grasp of the theory could plausibly explain our knowledge of the contents of the relevant indexicals in contexts.

Given these definitions, we can consider the following two principles:

[S&P+]   Our knowledge of the contents of simple indexicals of English in contexts is systematic and productive.

[C+]   If our knowledge of the contents of simple indexicals of English in contexts is productive and systematic, then there is a theory of character for English which is both adequate and internalizable.

These claims both seem quite plausible. But my main aim in this section has not been to defend them, but rather to defend the parallel between these claims and their non-+-ed counterparts. That is, I take myself to have argued for the following two conditional claims:

If [S&P] is true, then [S&P+] is true.
If [C] is true, then [C+] is true.

3 Demonstratives and intentions

Let's turn to the task of constructing a theory of character. Here and in what follows, I will focus on the special case of demonstratives—though much of the discussion will generalize to other sorts of indexicals.

My aim in this section will be to argue that, if there is any adequate internalizable theory of character for demonstratives to be had, it will be a theory which explains

the contents of demonstratives in contexts partly in terms of the intentions of the speaker of the context. The simplest intention theory is the following:

> *Simple Intention Theory*
> The content of a demonstrative d in a context c is the F iff the speaker of c intends that the F be the reference of d in c.

Complex intention theories agree with the Simple Intention Theory that speaker intentions to refer must be part of any internalizable adequate theory of the character of demonstratives, but deny that they are the whole story. So a complex intention theory will be of the form

> The content of a demonstrative d in a context c is the F iff
> (1) the speaker intends the F to be the reference of d in c, and
> (2) …
> …

There are a few features of this general sort of theory which might seem immediately objectionable. Two are dispensable, and are for present purposes distractions:

- One might object to the idea that the content of a demonstrative is the object for which the demonstrative stands—either on the grounds that one has a Fregean view of singular reference, or on the grounds that one takes demonstratives to be devices of quantification rather than genuine singular terms. These perspectives could easily be accommodated by modifying the foregoing to say 'some mode of presentation of the F' or by replacing talk of the content of the demonstrative with talk of the object which uniquely satisfies the demonstrative's restrictor. Since these sorts of worries will be orthogonal to the issues I want to discuss, I'll ignore them here and stick with the simple formulation above.
- One might object to the idea that the relevant speaker intentions are descriptive rather than de re. Surely sometimes our referential intentions are de re—but rather than complicating our formulations to make room for this, I'll simply treat the de re intention that o be the reference of d as the descriptive intention that the unique thing which has the property of being identical to o is the reference of d. This is just to ease some of the argument to follow; it could all be recast in a form which made the possibility of de re reference-fixing intentions explicit.

A different objection to the form of intention theories is not so easily skirted:

- One might object that intention theories require speakers to have intentions about objects being the reference of expression tokens. And while I have intentions of this sort, it's at least not obvious that my 3 year old daughter, who is a competent and demanding user of demonstratives, does.

It's not altogether obvious how to get around this problem, because it's not easy to find another content for the intention which will both do the needed reference-fixing work and which will be more plausibly attributed to my 3 year old. Indeed I think that this is a serious problem, not just for intention theories of demonstratives, but

also for attempts to use speaker intentions in describing the characters of comparative adjectives, expressions of quantification, and other context-sensitive expressions. But it is not the problem on which I want to focus here, so I will simply set it to the side. If in the end the problem proves insuperable it will, in a way which will soon be clear, be a help rather than a hindrance to my argument.

A first argument for an intention theory of the character of demonstratives begins by trying to think about how one might go about explaining the character of demonstratives without making use of speaker intentions at all. Here, I think, there are two main alternatives:

- Demonstration theories
- Salience theories

As we'll see, problems faced by these theories seem to require that, at crucial points, we appeal to the sorts of speaker intentions just discussed. Even if correct, this does not show that demonstrations and salience have no role to play in the character of demonstratives—it just shows that they can't do the reference-fixing work by themselves, and hence that their role is as part of a complex intention theory rather than an alternative to it. Indeed, this is something that I think that many theorists who give either demonstrations or salience a central role would be happy to accept. There is thus a sense in which the conclusion to be argued for in this section is fairly weak; the aim is not to show that intentions must take center stage in our theory of character, but only that they must play at least a supporting role.[7]

The simplest version of the demonstration theory identifies demonstrations with pointing gestures, and says that the character of a demonstrative is the function from contexts to objects which, for any context, delivers the object at which the speaker of the context is pointing—roughly, the first object met by a straight line which extends the finger or arm of the pointing gesture. Here are some of the main problems this sort of theory faces:

1. The theory is insufficiently general, because there appear to be uses of demonstratives accompanied by no demonstration (E.g., saying 'Wake that student up' with no accompanying pointing gesture when one student in a small classroom is openly snoring during discussion).
2. The theory can't handle cases of deferred reference, in which we manage to refer to someone by pointing at something else. King (2014) gives the example of saying 'he' while pointing at the chair of an absent student to refer to the student. One might try to argue that cases like this are outliers, and deserving of separate treatment; but, given the frequency with which they occur, this looks a bit ad hoc (Common examples: pointing at pictures to refer to the things pictured, pointing at maps to refer to locations).
3. As Kaplan (1970) notes, when we point at something we always point at many things (the person, his shirt, the button, …). Hence the character provided by our simple demonstration theory never seems to deliver a single object as the content of a demonstrative in a context.

---

[7] This an 'intention-sensitive semantics' in the sense of Stokke (2010).

4. Last, and perhaps most fundamentally, it is difficult to understand what a demonstration is without appeal to the intentions of the subject. It doesn't take long to see the sort of account of pointing gestures mentioned above is far too narrow; Gauker (2008) gives the nice example that "[h]iking in the mountains, I might point at wildflower using the shadow of my finger instead of my finger." This is surely right—but it is hard, I think, to explain why this counts as 'pointing at' the wildflower without bringing in the intentions of the pointer.

One might reasonably think that these cases just show that our version of the demonstration theory is too simple, and that we should simply look for a more sophisticated view of demonstrations. But it's hard to see what sort of view of demonstrations could help here—unless we treat intentions as essential to demonstrations, in which case we're simply pursuing an intention theory under a different name. That is less an objection to that sort of hybrid theory than it is simply to point out how broad the class of intention theories, as I've defined it, is— it is the class of theories which let intentions do some of the reference-fixing work, not the class of theories which let them do all the work. There is thus nothing to stop someone from pursuing a complex intention theory whose extra conditions make use of facts about demonstrations, and none of the points made so far show that a hybrid theory of this sort won't work. I'll return to this sort of complex intention theory below.

Note, further, that our Simple Intention Theory seems to give exactly the right result in each of the cases we've discussed. On that view, for example, there's no mystery about how I can refer to a student rather than the button on his shirt at which my finger is directed—this is explained just by my intending to refer to the former but not to the latter.

Let's turn to salience theories. The basic idea behind these theories is plausible enough: what a demonstrative in a context refers to has something to do with which objects are salient in that context. Pointing gestures, on this view, are just one way of making an object salient. In other cases, as in the case of the openly snoring student, the object is already salient, and no pointing gesture is required.

Intuitive as the theory is, it would be good to have some explanation of what salience amounts to. Here's how the notion is explicated on Allyson Mount's version of the view:

> By saying that something is salient, I mean that it is the focus of perceptual or cognitive attention. Salience, on this view, is not some objective feature that can be determined independently of the mental states of conversational participants; it is essentially a mind-dependent matter. An object is mutually recognized as maximally salient by conversational participants when all interlocutors have focused their attention on it, and are aware that they have all focused their attention on it. Thus my claim is that a demonstrative refers to the object mutually recognized as maximally salient. When there is no such object, the demonstrative does not refer. (Mount 2008, 154–5)

But this faces an obvious problem, which is that in the standard case there will be many objects on whom the conversational participants have focused their attention

(and mutually recognize each other as so doing). So, on Mount's account, there will typically be many objects which are maximally salient. Recall the example of pointing at the chair to refer to the absent student. It is plausible to say that the pointing gesture made the absent student the focus of cognitive attention. But it obviously also makes the chair the focus of cognitive (and perceptual) attention. So in virtue of what is the student, rather than the chair, the content of the demonstrative? This sort of salience theory gives us no answer.

It is overwhelmingly tempting, I think, to answer this question by saying: 'The speaker intended the demonstrative to refer to the student—that is why the student is the referent. Any audience, given the context, would be able to figure this out!' This is indeed plausible—but it is to give intentions a role in the determination of reference, and hence to move to a complex intention theory.

Nor can we limit the appeal to intentions to special cases. There will almost always be more than one object which is the focus of the perceptual or cognitive attention of the conversational participants, and is recognized to be so. The conversational participants themselves, for example, will typically satisfy this condition. Hence the role of intentions will be quite pervasive—and the prospects of a salience theory which does not make use of speaker intentions correspondingly dim.

One might try to get around this problem by saying that the content of a demonstrative is always the *most* salient object in the relevant context—the object on which the conversational participants are most intently focusing. But this leads to incorrect results. Suppose that you and I are trapped in a room with a tiger. The tiger will be, one would think, the most salient object in our conversational context. But, despite this, we will be able to use demonstratives to pick out other objects in the room—I might say, e.g., 'Let's make a run for that door.'

A second escape route is to appeal, not to facts about which objects are most salient, but to facts about salience raising. For consider again the tiger example. After I say 'that door' it is plausible that the tiger is still more salient to us than the door. But my utterance makes the door much more salient than it was before, whereas it does not increase the salience of the tiger. Perhaps, the salience theorist might say, this is the fact that explains why 'that door' refers to the door rather than the tiger: a given use of a demonstrative refers to the object whose salience is most increased by that very use of the demonstrative.

But this too will lead to incorrect results. If we are already completely focused on the tiger, my utterance of 'I wish that tiger were not in here' is unlikely to make the tiger more salient—but the demonstrative refers to the tiger nonetheless. But the more fundamental problem with the 'salience raising' theory is that it builds in just the facts that we hoped to explain. Presumably the utterance of 'that door' raises the salience of a particular door because the conversational participants recognize it as the reference of the demonstrative. But then we can hardly use the fact that the utterance raised the salience of that door as an explanation of the fact that it is the content, in that context, of 'that door.' This brings out the fact that the 'salience raising' theory of the character of 'that door' would be, even if adequate, not internalizable. We could not explain our knowledge of the contents of demonstratives in contexts in terms of knowledge of the rule that demonstratives in contexts always refer to the object whose

salience is most raised by their utterance, because in order to know which object satisfies that condition, we would first have to have mastered a rule which told us which object was the reference of the demonstrative in that context. And that is of course just what we want our theory of character to provide.[8]

So one argument for intention theories is that the competitors of such theories face problems which appear insuperable; and the most obvious responses to those problems involve building intentions into the story. Perhaps there are other theoretical alternatives we have not considered—but it's not easy for me to see what they would be.

A second argument in favor of intention theories of demonstratives is based on parallels with other indexical expressions. A plausible case can be made, for example, that comparative adjectives (e.g. 'tall') and quantifier expressions (e.g. 'all the beer') are also indexicals. But it is very hard to see how we could go about giving an account of the character of these expressions which did not make use of speaker intentions. The relevant comparison classes and domain restrictions are, after all, just not around to be demonstrated, so that sort of theory won't get off the ground. And one might think that if we have to appeal to intentions to handle some sorts of indexicals, we might as well appeal to them here.[9]

The conclusion of the argument of this section is the following:

> If there is an internalizable adequate theory of character for demonstratives in English, it is an intention theory.

4 The problem of conflicting intentions

The aim of this section is to point out a problem with the Simple Intention Theory. The basic problem is just that there are cases in which speakers have more than one intention with respect to the reference of a demonstrative, and these intentions can sometimes point in different directions. The classic example here is due to Kaplan (1970):

---

[8] See Gauker (2008) for a theory which makes use of facts about salience, but is quite different from the salience-based theories discussed above. I am inclined to think that Gauker's theory is either open to some of the counterexamples just discussed, or fails to be internalizable in the way that the 'salience raising' theory does, but I do not argue that here.

[9] Indeed, many of the problems to be discussed below would arise if we focused on the case of comparative adjectives, for example, rather than demonstratives.

One might worry that a theory of the character of demonstratives which was based wholly on the internal mental states of the speaker of the context would fail to recognize the importance of demonstrations in understanding these expressions—why would demonstrations so often be required if the character of the expression had nothing to do with those demonstrations? But to this question the intention theorist has a reasonably straightforward answer: in standard cases, the user of an indexical expression intends her audience to grasp the content of that indexical in the context. Hence she will intend her audience to be aware of whatever contextual features are relevant to the character of that indexical. If an intention theory is correct, then one such contextual feature is the speaker's own referential intention. So the speaker will intend, in the standard case, that her audience be aware of her intentions, and so will have a reason to make those intentions evident. Pointing gestures and the like are often the most straightforward way to do just that.

Carnap & Agnew
Suppose that without turning and looking I point to the place on my wall which has long been occupied by a picture of Rudolf Carnap and I say: [That] is a picture of one of the greatest philosophers of the twentieth century. But unbeknownst to me, someone has replaced my picture of Carnap with one of Spiro Agnew. …I have said of a picture of Spiro Agnew that it pictures one of the greatest philosophers of the twentieth century. And my speech and demonstration suggest no other natural interpretation …No matter how hard I intend Carnap's picture, I do not think it reasonable to call the content of my utterance true.

One might think that this shows, not just that the Simple Intention Theory is false, but that any intention theory is false. After all, all such theories say that speaker intentions provide necessary conditions on the contents of demonstratives in contexts—and isn't this just a case in which a speaker refers to something without having intended to refer to it?

But this overlooks the fact that the speaker in this scenario has more than one intention. He does of course intend to refer to the picture of Carnap; but he also intends to refer to the picture at which he is pointing. It's just that in this case the second intention wins out, and fixes the reference of the demonstrative. It is thus what I call a case of conflicting intentions.

Let's introduce a little bit of terminology which will make it easier to talk about cases of this sort. When a speaker intends a demonstrative in a context to refer to the *F*, I will say that the property of being the unique *F* is (or is one of) the speaker's *intended conditions*.[10] If a speaker, as in the case of Carnap and Agnew, has more than one intended condition, and one of those wins out by fixing the reference of the demonstrative, I will say that intended condition *trumps* the others (We can also talk interchangeably about intentions trumping other intentions—one intention will trump another iff its associated intended condition trumps the other associated intended conditions. In some cases it will be easiest to focus on the intentions, and in other cases on the intended conditions).

Though opinions differ about the right thing to say about the case of Carnap and Agnew, we can all agree that in that scenario the demonstrative does not refer to *both* the picture of Carnap and the picture of Agnew. This is all we need to see that the Simple Intention Theory is false, and that we need to move to a complex intention theory which adds at least one further condition on reference to the simple theory. Such a theory will be a theory of the form

The content of a demonstrative *d* in a context *c* is the *F* iff
(1) the speaker intends the *F* to be the reference of *d* in *c*, and
(2) for any other intended condition *G* such that the *F* ≠ the *G*, *F* trumps *G*
…

---

[10] As above, we can treat de re intentions to refer to some particular object *o* as the intention to refer to the unique thing which is identical to *o*.

Then what we need to fill this out is a theory of trumping: an account of the conditions (if any) under which one intention trumps another. Since speakers are able to determine the referents of demonstratives in contexts even in cases of conflicting intentions, we can conclude the following:

> If there is an internalizable adequate intention theory of character for demonstratives in English, there is an internalizable adequate theory of trumping.

where an adequate theory of trumping is one which delivers the correct results about when one intention trumps another, and an internalizable theory of trumping is one which might plausibly explain our ability to tell when one intention trumps another.[11]

One might push back here against the claim that our theory of trumping must be internalizable, on the grounds that hearers cannot read the speaker's mind and hence don't face the problem of conflicting intentions at all.[12] Against this, there are two main things to be said.

The first is that hearers do often face problems of conflicting intentions, because they do often know things about the speaker's intentions. In the Carnap/Agnew case, for example, the audience presumably knows that the speaker intends to refer to the picture at which he is pointing, and also (since the speaker cannot be supposed to think that Agnew is a great philosopher) that the speaker intended to point at some picture other than the one of Agnew. So the audience is in a position to see that the speaker's referential intentions conflict (even if they would be unlikely to express it in quite those terms). Despite this, they would be able to assign the picture at which the speaker is pointing as the reference of the demonstrative in that context. Since competent language users can do this in novel contexts, one can then argue (in a by-now familiar way) that their ability to identify the referent must be explained by their knowledge of an internalizable adequate theory of character.

The second is that when speakers consider thought experiments involving imaginary uses of demonstratives (about which more in Sect. 5 below), in which they are given explicit statements of the (imaginary) speaker's intentions, they have very clear intuitions about the content of the expression in the imagined context. But we can then ask how competent speakers are able to identify the referent in these imaginary contexts; and since we can multiply these contexts indefinitely, this again seems best explained by their knowledge of an internalizable adequate theory of character.

Let's now turn to the question of how we might give an internalizable adequate theory of trumping.

---

[11] It is again worth noting that there is nothing very special here about demonstratives; in the case of any context-sensitive expression whose character is at least in part explained in terms of speaker intentions, we will get cases of conflicting intentions. One might even think that examples like those discussed in Romdenh-Romluc (2002) suggest that similar problems can arise for 'I.'

[12] Thanks to an anonymous referee for pressing this point.

5 Theories of trumping

I think that there are broadly three ways to go about giving a theory of trumping. The simplest one is what I call a 'no tolerance' theory—on this sort of view, the existence of conflicting intentions entails reference failure, and hence no intentions ever actually trump other intentions. The second sort is what I call a 'special intention theory,' which tries to identify some special sort of content which is such that intentions with that kind of content always trump the others. The last is what I call a 'relational theory'—on this sort of theory, we try to explain which intentions trump which in a certain context by looking at the relations in which those intentions happen to stand in a particular context. This last sort of theory is by far the most plausible, I think; I'll briefly explain the problems with the first two sorts of view before moving on to discuss some versions of the third.

5.1 No tolerance theories

King (2014) was inclined toward the no tolerance view, which we might express as the following very simple theory of trumping:

(A)   Nothing trumps anything

This delivers the result that the speaker in Carnap and Agnew fails to refer to anything, since, given (A), neither intention will satisfy condition (2) of the formulation above.

This strikes me as a quite counterintuitive claim. But a case can be made for it; one might argue that the speaker in that context is so confused about his environment that his attempt to refer to something simply misfires.[13] But other examples of conflicting intentions show, I think, that the no tolerance theory simply cannot be right. Reimer (1991) presents a case which seems to involve conflicting intentions, and which seems to involve genuine reference:

> Suppose, for instance, that I suddenly realize that I have left my keys on the desk my (shared) office. I return to my office, where I find the desk occupied by my officemate. I then spot my keys, sitting there on the desk, alongside my officemate's keys. I then make a grab for my keys, saying just as I mistakenly grab my officemate's keys, "These are mine."

It is quite difficult to deny, I think, that in this case 'these' refers to the keys mistakenly grabbed. This is true despite the fact that the speaker intends both to refer to her own keys and to the keys she grabs—and that these intentions conflict.[14]

---

[13] See King (2013, 10–11) for an interesting argument that 'says' and 'agrees' ascriptions support the verdict of reference failure. While this is suggestive, I think that those tests yield different results in other cases of conflicting intentions, like the case of the carnival discussed below.

[14] Reimer (1991) presents this case not as one where one intention trumps another, but as one where the demonstration trumps speaker intentions. But in Reimer (1992), Sect. 3, she is clear that the case of the

Here's another case which is perhaps even clearer on this score:

The carnival
I'm at a carnival where, for a fee, I can choose from among a large number of plastic balls, one of which contains $100. The person running the game asks me which one I want and, pointing, I say, 'That ball.' I intend to refer to the ball at which I am pointing (say, ball #58), and of course also intend to refer to the ball which contains $100. But it turns out that another ball – ball #113 – contains the cash. The intentions conflict – but it is clear that the content of my utterance of 'that ball' is ball #58.[15]

These cases show that we need a theory of trumping which sometimes actually delivers the result that one intended condition trumps another.

5.2 Special intention theories

The simplest way to provide such a theory would be to single out a class of always-trumping intentions. But which class?

5.2.1 De re intentions

When thinking about the case of the carnival just described, one might think that the relevant difference is that the intended condition associated with my pointing gesture is de re, whereas the other is merely descriptive ('the winning ball, whatever it is'). This might suggest:

(*B*)   De re intended conditions always trump all of the others.

But we can imagine a variant of the above case in which the speaker has a prior acquaintance with the winning ball, and a de re intention to refer to it. Perhaps, upon his prior acquaintance, the speaker named the ball 'Larry.' He will then have the de re intention to refer to Larry, and we would then have a case of conflicting de re intentions. Both of them can't trump the other. Further, and even worse, in the original Carnap/Agnew case, we seem to have a descriptive intended condition trumping a de re intended condition.

Footnote 14 continued

keys is a case of conflicting intentions. (In her terms, it is a conflict between primary and secondary intentions.) For a discussion of the positive theory of Reimer (1992), see Sect. 5.2.2 below.

[15] King (2013) (notes 15 and 28) doubts whether in this scenario the speaker really intends to refer to the winning ball, noting, plausibly, that it's not clear that people buying lottery tickets can intend to buy the winning ticket. But it is not hard to modify the case to avoid this objection: we can just imagine that the person playing the game has received a false but plausible anonymous tip about which ball has the cash in it. In these conditions the subject can surely rationally believe that he will refer to the winning ball, intend to refer to the winning ball, and this intention can be part of various of the subjects plans in the sense required by King's view of intentions. For more on King's take on these cases, see Sect. 5.3.3 below.

5.2.2 Demonstrative intentions

But perhaps the category of de re intentions is simply the wrong choice for the class of always-trumping intentions. One notable feature of the cases of conflicting intentions discussed so far is that demonstrations seem to play some special role; intentions to refer to thing one is demonstrating seem always to trump other sorts of referential intentions. This seems to explain our intuitions in the cases of Carnap and Agnew, the carnival, and Reimer's example of the keys.

Hence, adapting the view of Reimer (1992), herself adapting the view of Kaplan (1989), we might hold that the intentions which accompany demonstrations play a special role here; perhaps the trumping intention is always, in Reimer's sense, a secondary intention. In the case of Carnap and Agnew, this would be "an intention to demonstrate, and say something of, the picture in the direction of the gesture" (390). More generally, on this view, the trumping intention will be the intention to refer to the object demonstrated. This view might be stated as follows:

(C) Demonstrative intended conditions always trump all of the others.

This proposal seems initially to be open to an objection which can be brought out by the following case:

> *Incompetent pointing*
> A student in the front row of Philosophy 101 is openly napping. I point at him, saying 'Somebody wake that student up.' I intend 'that student' to demonstrate the sleeping student, and intend 'that student' to demonstrate the thing at which I am pointing. But in fact my finger is aimed just over the student's left shoulder, in such a way that a line drawn in the direction indicated by my finger would intersect a student sitting in the second row.

My demonstrative intention appears to be the intention to refer to the thing at which I am pointing. But in this case that demonstrative intention appears to be trumped by the intention to refer to the sleeping student.

Here I think that the proponent of (C) has a reasonable reply: she can object that this objection misconstrues the nature of demonstrative intentions. If we typically did intend to refer to whatever we were pointing to, we would be much more careful with our pointing gestures than we in fact are. And, correspondingly, if the teacher really were relying on the students being able to see exactly which student he was pointing at, he would be radically overestimating their ability to figure out the angle of his finger relative to the seated students. Rather, I think, in cases like this the relevant demonstrative intention is something like the intention to refer to the sleeping student in the general direction of the teacher's pointing. And if this is the relevant demonstrative intention, cases like incompetent pointing pose no problem for the present view.

This brings out an important fact about demonstrative intentions, to which I will return below. They are not plausibly thought of as the intention to, for any x at which I am pointing, refer to x. This is shown by the case of incompetent pointing

just discussed, but is also shown by a point made above in our discussion of demonstration theories: whenever I point at one thing, I point to many things. But I do not intend to refer to all of those things via a demonstrative whose use is accompanied by the relevant pointing gesture.

This shows that demonstrative intentions are always, for some property *F*, the intention to refer to the *F* in the general direction of my pointing gesture. Then our question is: which property?

The answer to that question suggested above was: the property of being a sleeping student. But a moment's reflection shows that this answer gives us the incorrect result in some cases. Just imagine that the student who appeared to be sleeping was merely pretending to nap, as a way of protesting the less than exciting lecture. 'That student' would still refer to the faux-sleeping student, despite the fact that student would not be the sleeping student in the rough vicinity of my pointing. This would then be a case in which, contra (*C*), the demonstrative intention is trumped.

One might reply by saying that this is not the only demonstrative intention that the speaker has: he also intends to refer to the student who *appears to be sleeping* in the vicinity of his pointing gesture. And that choice of a property to figure in the demonstrative intention seems to give us just the result we want.

But this is less a defense of (*C*) than an admission of its failure to solve the problem that we're interested in. For the fact that the speaker intends to refer to the student who appears to be sleeping in the vicinity of his pointing gesture does not change the fact that the speaker also intends to refer to the sleeping student in the vicinity of his pointing gesture. And of course these two intentions could conflict. Just imagine that next to the student pretending to be sleeping, there is another student cleverly sleeping with eyes wide open and an attentive look on his face. One demonstrative intention singles out the first student, and another demonstrative intention singles out the other. The first of these trumps the second. But what determines when one demonstrative intention trumps another? This is the question to which we need an answer, and (*C*) does not provide it.

5.2.3 Referential intentions

Bach ([1992](#)) suggests instead that

> The relevant intention …the specifically referential one, is part of a communicative intention, a reflexive intention whose distinctive feature is that 'its fulfillment consists in its recognition' by one's audience, partly by supposing (in Gricean fashion) that the speaker intends his intention to be recognized. …a referential intention isn't just any intention to refer to something one has in mind but is the intention that one's audience identify, and take themselves to be intended to identify, a certain item as the referent by means of thinking of it a certain identifiable way.

This suggests:

(*D*)   Referential intended conditions always trump all of the others.

But how should we apply this to the carnival case? I do, let's suppose, have the following intention:

> (i) I intend that my audience identify ball #58 as the referent of 'that ball' on the basis of their recognition of my intention.

If (i) is the speaker's only referential intention, then (*D*) delivers the wanted result that ball #58 is the content of the indexical in the context.

But let us suppose that the speaker has good reason to believe that he knows which ball is the winning ball, and that his audience knows this. Then he will also intend to refer to the winning ball, and intend that his audience recognize the winning ball to be the referent of 'that ball' on the basis of their recognition of his intention. So the speaker may have the following intention:

> (ii) I intend that my audience identify the winning ball as the referent of 'that ball' on the basis of their recognition of my intention.

This appears to be a referential intention, but conflicts with intention (i). In this case, (i) trumps (ii)—but if both are referential, then this fact cannot be explained by (*D*).

One might reply by saying that intention (ii), being descriptive, cannot be a genuinely referential intention. But the case is easily modified to avoid this objection. For simply imagine, as above, that the speaker has prior acquaintance with the winning ball, which he calls 'Larry.' Then the speaker will have the following intention:

> (iii) I intend that my audience identify Larry as the referent of 'that ball' on the basis of their recognition of my intention.

And (i) and (iii) are conflicting intentions which are, in Bach's sense, referential. But if we can have conflicting referential intentions, we can't solve the problem of conflicting intentions by saying that referential intentions always trump.

Indeed, there is another route to the claim that speakers can have both intentions (i) and (iii) which does not rely on the assumption that the speaker believes, and the audience knows the speaker believes, that the speaker will be able to select the winning ball. For the speaker may simply be both wildly and irrationally overconfident, and a convinced Millian. Then the speaker will (mistakenly, as it turns out), take intention (i) to be identical to intention (iii), and thus form the conflicting referential intention (iii). Here again intention (i) trumps—but not in virtue of being a referential intention.

5.3 Relational theories

It looks tough to find a special class of always-trumping intentions. Better, perhaps, to look to the relations which hold between the relevant intentions in particular contexts.

5.3.1 Salience

One idea is that we could bring the resources of salience theories of demonstrative reference to bear on the present problem for intention theories. The basic idea would

be that, in cases of conflicting intentions, the trumping intended condition is the one which is satisfied by an object which is salient in that context.

Let's see how this looks if we apply Mount's account of salience:

(E) *F* trumps *G* iff the *F* is the focus of the cognitive attention of the conversation participants (and mutually known to be so) and the *G* is not.

Given the foregoing, one problem here is probably obvious: there can be cases of conflicting intentions which are such that the objects of both intended conditions are salient in the conversation. The version of the example of the carnival in which there is prior acquaintance with the winning ball is such a case, and the example of Carnap and Agnew can be turned into one if we imagine that the picture of Carnap is not gone from the room, but simply moved to a different location to the right of the speaker.

A second problem is that this will give the wrong result in cases in which both the speaker and the audience have false beliefs about their environment. Here is such a case:

Mutual hallucination
You and I are having a conversation while jointly experiencing a quite convincing mutual hallucination of your cat 'Fluffy.' Pointing at what I take to be the Fluffy, I say 'That cat …'

This may not at first seem like a case of conflicting intentions, but it is. I intend to refer to Fluffy, but also intend to refer to the cat at which I am pointing. The first intention singles out Fluffy; the second intention singles out nothing. Given that Fluffy is the object of our cognitive attention, (E) dictates that the first intention trumps. But, intuitively, this is a case of reference failure, which means that the second intention (whose intended condition picks out nothing) trumps.

A third problem—which is also a problem with simple salience theories—is that this makes the conditions on demonstrative reference much too strong. For suppose that my audience is napping, or drunk, or distracted—they might well not be focusing on the intended referent of the demonstrative. But this bad luck on my part is surely not sufficient for reference failure.

5.3.2 Audience uptake theories

One might try to get around all three of these problems by modifying (E) in certain ways. In response to the first problem, we might require not just that the object be salient to the conversational participants, but also that the conversational participants identify the object as the content of the demonstrative in the context. In response to the second and third problems, we might focus less on what the actual audience does than on what an idealized audience would do.[16]

---

[16] The second and third problems for salience theories seem to me to be problems for the audience uptake condition endorsed in Stokke (2010).

Incorporating both modifications gives us

(F) The *F* trumps the *G* iff were the conversational participants ideal (i.e., attentive, competent with the language, not experiencing hallucinations, …), they would identify the *F* as the content of *d* in *c*.

But, given our aims, this faces a decisive objection. Even if this theory of trumping is adequate, it is not internalizable. Suppose that I am trying to teach someone how to interpret 'that' in arbitrary contexts. I can hardly help them along by telling them that they should take 'that' to have the content which an ideal interpreter would take it to have. I'd have to, in addition, give them the rule which an ideal interpreter would follow. But a rule of this sort is exactly what (F) was supposed to provide.[17]

5.3.3 Explanatory relations

A different and more promising approach begins with the thought that a subject's intentions are usually not one-off things, but rather are structured together into plans.[18] In the standard case, we have some of our intentions because we have certain other intentions, along with other mental states. So, for example, in the case of Carnap and Agnew, the speaker intends to refer to the picture of Carnap. Because he has this intention, plus the (false) belief that this picture is behind him, he forms the intention to point at, and refer to, the picture behind him. We might represent this explanatory chain as follows:

> intention to refer to the picture of Carnap (+ beliefs that the picture of Carnap is behind him) ⤳ intention to refer to the picture behind him

And, in general, the user of a demonstrative who has several referential intentions will be such that those intentions are structured into a plan of the following sort:

> intention to refer to the *F* (+ some other mental states) ⤳ intention to refer to the *G* (+ some other mental states) ⤳ …

But this chain will come to an end at some point, and there will presumably be some last referential intention which stands in the closest relation to the actual utterance of the demonstrative in the context. Following King (2013), call this intention, the one that comes last in the structured chain of intentions, the *controlling intention*. Call the intended condition of the controlling intention the *controlling intended condition*. King suggests the following view:[19]

(G) Controlling intended conditions always trump all the other intended conditions.

---

[17] In the end, I think that (F) is probably inadequate as well as non-internalizable, for reasons discussed in Speaks (forthcoming).

[18] See, among others, Bratman (1987).

[19] He doesn't endorse the view, however; he offers it as the best view for someone who thinks that the demonstrative in Carnap/Agnew refers to the picture of Agnew, rather than being an instance of reference failure—while leaving it open which is the correct view of that, and other similar, cases.

As we've already seen, this seems to give us the result we want in the case of Carnap and Agnew. It also seems, for much the same reason, to give us the right result in the case of the carnival. Indeed, I think that something like (G) is the most plausible treatment of conflicting intentions on offer. But in the end, I think that the view fails for reasons somewhat related to those given above for rejecting (C): sometimes, in cases where a speaker is mistaken about the properties of some object in her environment, her controlling intention will single out an object other than the referent of the relevant demonstrative.

Recall the example of the professor who sees a student who appears to be sleeping, intends to refer to the sleeping student, and executes this intention with an utterance of 'Wake that student up.' Here, it seems, the speaker's controlling intention is the intention to refer to the sleeping student. But we have already seen, in our discussion of (C), that this demonstrative intention does not fix the reference of the demonstrative, since the demonstrative refers to the student even in cases where the student is not really sleeping, and hence does not satisfy the relevant intended condition.

I can think of two ways for the proponent of (G) to handle this sort of case—both of which are ways of avoiding the conclusion that the intention to refer to the sleeping student really is the speaker's controlling intention in this case.

The first line of response begins by pointing out that the speaker in this case has, not just the purely descriptive intention to refer to the sleeping student, but also the de re intention to refer to *that student*—the one who appears to be sleeping. And on one plausible construal of the case, it is really this de re intention which is the professor's controlling intention. If S is the apparently sleeping student, we might represent the plan leading to the professor's utterance as follows:

standing intention to refer to any sleeping student ($+$ belief that S is sleeping)
⇝ intention to refer to S

Here the de re intention to refer to S, rather than the intention to refer to the sleeping student, comes last in the explanatory chain, and hence (on this view) fixes the reference of 'that student.' And this gives us the result we want, whether or not the student is really sleeping.

The problem with this line of response is not that it is incorrect, but that it is insufficiently general. The professor's plan may be structured in the way just described; but it also may not be. The professor may also begin with the intention to refer to S, and then, upon turning his attention to S, form the false belief that S is sleeping—which in turn gives rise to the intention to refer to the sleeping student. The professor's plan in this case could be represented as follows:

intention to refer to S ($+$ belief that S is sleeping) ⇝ intention to refer to the sleeping student

In this version of the story, the intention to refer to the sleeping student is the controlling intention; but the demonstrative still refers to S, even in the case where S is not sleeping.

This version of the case is exactly parallel to the case of Carnap and Agnew. In each case, we begin with a de re intention (to refer to the picture of Carnap, to refer to S); in

each case, the speaker has a false belief about the object of that de re intention (that the picture is behind him, that S is sleeping); and in each case, this gives rise to a descriptive controlling intention (to refer to the picture behind him, to refer to the sleeping student). The problem for the proponent of (G) is that whereas this descriptive controlling intention plausibly fixes the reference of the demonstrative in the case of Carnap and Agnew, it doesn't in the case of the seeming-to-be-asleep student.

This parallel between the cases heads off some possible moves that the proponent of (G) might make. It would be hard, for example, to deny that the intention to refer to the sleeping student is the controlling intention in this case without also denying that the intention to refer to the picture behind him is the controlling intention in the case of Carnap and Agnew. In both cases, after all, the intention informs the way in which the speaker carries out his plan to refer to a specific object. We can imagine that the professor's audience would have been unable to figure out the reference of 'that student' had the student not appeared to be sleeping—just as the audience in Carnap/Agnew would have been unable to figure out the reference of 'that' without the speaker's pointing gesture.[20]

A second line of response focuses not on the professor's de re intention, but rather on the fact that the professor in this case plausibly intends, not just to refer to the sleeping student, but also to refer to the student who appears to be sleeping.[21] Of course, the fact that the professor simply has this intention is not enough; that intention must also figure in the professor's referential plan, and it must figure as the controlling intention, rather than an intermediate intention, in that plan. Is it plausible that it does so figure?

In answering this question, we'll have to be a bit clearer about the phrase 'appears to be sleeping.' We should distinguish the following two properties:

(i)   the property of appearing to everyone in the class to be sleeping
(ii) the property of appearing to the professor to be sleeping

Here is a plausible just-so story that might seem to give the proponent of (G) what she wants. The professor intends that the student he takes to be sleeping be the reference of his use of a demonstrative. He further wants his class to understand who the reference of the utterance of the demonstrative is. The professor forms the belief that the student is not just sleeping, but is obviously sleeping—so that it is presumably evident to those around him that he is sleeping. This gives the professor a way to execute his plan, since a pointing gesture will make plain the intended

---

[20] Another move would be to say that the controlling intention in the case of the professor is not a purely descriptive intention, but rather a descriptively enriched de re intention to refer to S, the sleeping student, and that in the cases of such controlling intentions, the reference of the demonstrative is fixed by the singular element rather than the extra descriptive material. But the problem again is why, if we say this in the case of the professor, we should not say this in the case of Carnap and Agnew—which would give us the incorrect result that the demonstrative in that case refers to the picture of Carnap.

[21] One might object here that, while the professor may have had this extra intention, he may not have. I can intend to refer to the F without intending to refer to something which appears to be F. Otherwise, I would have to intend not just to refer to the thing which appears to be a sleeping student, but also to the thing which appears to appear to be a sleeping student. But it is no doubt true that in the natural way of understanding the case of the sleeping student, the professor does have an appears-intention of some sort.

reference to his audience. So he forms the intention to refer to the student whom everyone in the class does, or can, see to be sleeping.

Note that this just-so story relies essentially on property (i), rather than property (ii), being the controlling intended condition. The fact that the student appears to *him* to be sleeping hardly gives him the means to execute his plan of getting the class to see that the relevant student is the reference of his utterance of the demonstrative. And indeed no just-so psychological story of the relevant sort would appear to be available for property (ii).

The problem is that if we take property (i) to be the controlling intended condition, we can simply recreate a version of the problem which besets the view that being the sleeping student is the controlling intended condition: we can just imagine a version of the case where, despite the fact that the student appears to the professor to be sleeping, the student does not appear to everyone in the class to be sleeping. And this does not have to be because the students are distracted, blind, or in some other sense less than ideal observers—though those cases too are problematic. It might be simply that the professor has misperceived the student; the professor thought that the student appears to be sleeping [in sense (i)], but the student does not. Or (a more far out version of the case) it might be that the students, who as it turns out are quite sophisticated, have constructed an elaborate hoax using holograms designed to produce the result that the student appears to be sleeping only from the vantage point of the professor's lectern.

I think that the problem posed by these cases is a general one, and is similar in some ways the problems of misdescription emphasized in Kripke's 'semantic argument' against descriptivist theories of the reference of names. Cases like the fake-sleeping student push the proponent of (G) to find controlling intended conditions which are such that the speaker cannot be mistaken about whether the referent of the demonstrative satisfies that condition. But once we do that, the intended conditions we find—like property (ii) above—simply can't play the role in the psychologies of speakers that they would have to in order to be parts of the contents of controlling intentions.

So far I've considered some prima facie plausible attempts to provide an adequate and internalizable theory of trumping, and argued that none is adequate save, perhaps, (F), which is non-internalizable. Of course, to show that a bunch of theories fail is not to show that there is no true theory of the relevant sort that we haven't been smart enough to think of yet. But I do think that the examples discussed so far show that if there is an adequate theory of trumping to be had, it will not be very simple or natural. And given the theoretical enterprise in which we're engaged, one might think that the complexity of a theory of trumping is a point against it. After all, given that we are looking for a theory which is not just adequate but internalizable, our account of the character of demonstratives must not just get the cases right but also play a real psychological role in helping us to identify the contents of demonstratives in various contexts of use. And it is perhaps not especially plausible that a theory of character which is excessively complex, gerrymandered, or unnatural plays such a psychological role.

For this reason, the generality and difficulty of the problem of conflicting intentions makes a strong case for:

There is no internalizable adequate theory of trumping.

6 The problem

The problem, by this point, is pretty clear. We've seen that an argument can be made for each of the following seven theses:

(1)   [S&P] is true.
(2)   [C] is true.
(3)   If [S&P] is true, then [S&P+] is true.
(4)   If [C] is true, then [C+] is true.
(5)   If there is an internalizable adequate theory of character for demonstratives in English, it is an intention theory.
(6)   If there is an internalizable adequate intention theory of character for demonstratives in English, there is an internalizable adequate theory of trumping.
(7)   There is no internalizable adequate theory of trumping.

(1)–(7) are jointly inconsistent, so at least one must go. But it is very hard to see how any of them could be false. To recap:

- If (1) or (2) is false, then virtually all of contemporary semantics is based on a mistake—and our ability to understand novel sentences is a mystery. This would be an interesting result—but also one which is pretty hard to accept.
- It is hard to see how (3) or (4) could be reasonably rejected, given that our knowledge of the contents of indexicals in arbitrary contexts seems to be parallel in the ways detailed above to our knowledge of complex expressions.
- (5) seems plausible just because alternatives to intention theories seem either incomplete or implausible.
- Given the existence of cases of conflicting intentions which are not cases of reference failure, (6) must be true.
- If one rejects (7), one owes a solution to the problems discussed in Sect.5, which seem to me pretty difficult.

I hope that it is now clear why I framed the discussion of demonstrative reference in the way that I did. Once we see the parallels between novel utterances and our understanding of demonstratives, we have the resources to challenge two widely held views about demonstratives.

The first—which I have found is a quite common response to the problem of conflicting intentions—is a view which simply denies that demonstratives have semantic contents relative to contexts, and hence denies that there is any true theory of the character of demonstratives to be given.[22] Proponents of this sort of view deny that the character/content model which seems to fit 'I' so neatly can be

---

[22] Thanks to Kent Bach and Scott Soames for helpful discussion of this sort of view. For versions of this kind of view, see Higginbotham (2002), Bach (2005, 2006, ms).

extended to demonstratives, and hence deny that sentences involving demonstratives have truth conditions relative to contexts.

One might have thought that, by endorsing this sort of view, one could escape the problem of conflicting intentions. After all, if one denies that demonstratives have a character, one hardly owes an internalizable adequate theory of their character. The present way of situating the problem shows that this is a mistake. For, as we've seen, speakers are able to determine the content of a demonstrative in novel contexts; their ability to determine the content of a demonstrative is productive. Either we must explain this ability in terms of knowledge of a theory which tells us the content of a demonstrative relative to a context, or we do not. If we take the former option, we need an internalizable adequate theory of trumping. If the latter, it is then mysterious why we would need to explain the productivity of our general semantic competence via knowledge of a compositional semantic theory.

It is only a verbal solution to reply to this by saying that speakers don't really have the ability to determine the *content* of demonstratives relative to contexts, but only the ability to determine the *speaker reference* of a demonstrative in various contexts. For let the *schmaracter* of an expression be a function from contexts to the speaker reference of that expression in the context.[23] Then we can ask whether the ability of competent speakers to identify the speaker reference of a demonstrative in a context is to be explained in terms of their knowledge of an internalizable adequate theory of schmaracter, or not. If it is, then we are still owed the internalizable adequate theory of trumping which we do not know how to provide. If it is not, then it is again mysterious why the productivity of our knowledge of speaker reference is not to be explained in the same way as the productivity of our semantic competence.

Someone who wants to escape our problem by denying that demonstratives have a character must therefore do one of two things: either find some disanalogy between our competence with novel sentences and our competence with novel indexical/context pairs which would make plausible the denial of (3) or (4), or give up on the standard Fregean argument in favor of the existence of a compositional semantic theory [thus denying (1) or (2)]. Neither looks like an especially easy way to go.

The second view which is challenged by this way of situating questions about the character of demonstratives is the view that character is to be explained, at least in part, in terms of what an audience does or would do under certain conditions. On this sort of approach, the very existence of the intuitions about demonstrative reference which give rise to the problem of conflicting intentions can be used to solve that problem. But theories of this sort are, even if adequate, obviously not internalizable. There must be some way for hearers to identify the content of a demonstrative in a context without relying on facts about what other hearers do or would do. Hence the proponent of this sort of theory must either explain why our understanding of novel utterances is not best explained in terms of the existence of

---

[23] Here I am taking 'context' in a sufficiently broad way—as, e.g., a centered world—to ensure that there will be such a function.

an internalizable compositional semantic theory, or, again, find some disanalogy between our understanding of novel utterances and our understanding of novel indexical/context pairs which makes the former but not the latter apt for explanation in terms of grasp of an internalizable theory.

The moral of the story is not, of course, that there is no way out of the puzzle. (1)–(7) are jointly inconsistent, so at least one must be false. The optimistic thought is surely that it is (7) which is to blame. And that is indeed the one that I incline toward rejecting. But, as I have tried to show, the problem of conflicting intentions, while not new, is much, much harder than is ordinarily supposed. And, unfortunately, it is a problem which is difficult for anyone—at least, anyone who shares the orthodox Fregean motivation for compositional semantics—to simply sidestep.

References

Almog, J., Perry, J., & Wettstein, H. (Eds.). (1989). *Themes from Kaplan*. Oxford: Oxford University Press.

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

Bach, K. (2005). Context ex machina. In Z. G. Szabó (Ed.), *Semantics versus pragmatics* (pp. 15–44). Oxford: Oxford University Press.

Bach, K. (2006). What does it take to refer? In E. Lepore & B. Smith (Eds.), *The Oxford handbook of philosophy of language* (pp. 516–554). Oxford: Oxford University Press.

Bach, K. (ms). Reference, intention, and context: Do demonstratives really refer?

Bratman, M. (1987). *Intention, plans, and practical reason*. Center for the Study of Language and Information.

Frege, G. (1914/1980). Letter to Jourdain. In G. Gabriel et al. (Ed.), *Philosophical and mathematical correspondence*. Chicago: University of Chicago Press.

Gauker, C. (2008). Zero tolerance for pragmatics. *Synthese*, *165*, 359–371.

Higginbotham, J. (2002). Competence with demonstratives. *Philosophical Perspectives*, *16*, 1–16.

Kaplan, D. (1970). Dthat. *Syntax and Semantics*, *9*, 221–43.

Kaplan, D. (1977/1989). Demonstratives. In J. Almog et al. (1989) (pp. 481–563).

Kaplan, D. (1989). Afterthoughts. In J. Almog et al. (1989) (pp. 565–614).

King, J. C. (2013). Supplementives, the coordination account, and conflicting intentions. *Philosophical Perspectives*, *27*, 288–311.

King, J. C. (2014). Speaker intentions in context. *Noûs*, *48*, 219–237.

Mount, A. (2008). Intentions, gestures, and salience in ordinary and deferred demonstrative reference. *Mind and Language*, *23*, 145–164.

Reimer, M. (1991). Demonstratives, demonstrations, and demonstrata. *Philosophical Studies*, *63*, 187–202.

Reimer, M. (1992). Three views of demonstrative reference. *Synthese*, *93*, 373–402.

Romdenh-Romluc, K. (2002). Now the French are invading England!. *Analysis*, *62*, 34–41.

Speaks, J. (forthcoming). The roles of speaker and hearer in the character of demonstratives. *Mind*.

Stokke, A. (2010). Intention-sensitive semantics. *Synthese*, *175*, 383–404.