STANDING IN THE WAY OF A SCIENCE OF MEANING:
MAINSTREAM SEMANTICS + DEFLATIONARY TRUTH

Presenter: Dr. Alexis Burgess, Stanford University

Thanks very much for that introduction. I'm very happy to be here. I'm a little bit worried though, that I might have chosen the wrong topic for today, so I'll try to work in genocide or…

[Background laughing]

I really am afraid that this is going to be terribly dull, but I'll try to make it as lively as possible. And I have to apologize, I usually don't just read papers, but I found myself in the position where if I don't do it I'm just going to get things wrong. So I'm just going to read this thing and make as much eye contact as possible. And a special apology for those of you who took the time to read the paper beforehand. So for those people I'll make a special effort to insert italics, because they don't let us do that in the philosophy journals anymore.

Okay, so the bulk of contemporary semantics, a theory of meaning, aims to explain systematically,why sentences have the truth conditions and stand in the implication relations that they seem to. So intuitions about the truth or falsity of sentences and validity or invalidity of inferences in context are used to justify hypotheses about the meanings or semantic values of the significant parts of sentences, in accordance with some version of the vague, but venerable principle of compositionality. So such hypotheses are then offered up as theoretical explanations of the intuitive data about truth conditions and inferential relations in much the same way that hypotheses in chemistry, say, are invoked to explain macroscopic phenomena. So properties of and relations between the parts account for features of the whole.

So this general approach to semantics is widely thought to be fundamentally at odds with deflationism, which is a family of theories about the meaning or expressive role of the phrase "true is true," which is often attended by effacing metaphysical claims about the nature of truth itself. Well, I guess we can just say at this point, according to the deflationists there's not much more to say about truth once you've noted that all instances of something like this equivalence are true. In particular, "snow is white" is true if and only if snow is white. More to say about deflationism later.

But anyway, as the point is sometimes put, mainstream formal semantics seems to presuppose some version of the competing correspondence theory of truth. So deflationism, therefore, in the eyes of many people, stands in the way of our developing science of meaning - just as philosophical qualms about abstract objects like numbers, might foolishly be taken to stand in the way of mathematics. So this apparent tension between mainstream semantics and deflationary theories of truth, has led to a somewhat acrimonious polarization of the philosophical study of meaning. In print and in person, practitioners of truth conditional or model theoretic semantics routinely accused deflationists of willful blindness for the countless explanatory successes that have stemmed from their formal framework. And deflationists routinely complain that the philosophical commitments implicit in that framework, have been inadequately scrutinized by technocrats and linguistics.

[Background laughing].

You can get a really nice feel for this sociological situation by Googling Jason Stanley later blog, "Deflationary Use Theories of Meaning." There's an extremely entertaining exchange between luminaries in the field who just seem to be talking past each other for pages and pages. Okay, so skip that.

The various use theories of meaning and inferentialisms, quote unquote, have been enlisted to counter the charge that truth conditional semantics is the only game in town, but critics generally remain unimpressed by the rigor and explanatory power of such alternatives. The debate's been complicated by related philosophical disagreements over the status of compositionality and semantics, something called semantic externalism, and the alleged normativity of meaning, just to get the list of neighboring battlegrounds started. But if there's one thing most parties in this debate will agree to, it's that either there is something seriously wrong with deflationism or there's something seriously wrong with contemporary mainstream model theoretic semantics. So the present paper is an attempt to flush out why I think this is just wrong. So it seems to me that certain kinds of deflationism about the truth are perfectly compatible with mainstream model theoretic semantics.

Alright, so here's the plan for the talk. Section one, I'm going to characterize deflationism, mainstream semantics, and the correspondence theory of truth in as much a detail as should be required for the subsequent arguments. Section two, I'm pretty quickly going to try to undermine three arguments for the incompatibility of deflationism in mainstream model theoretic semantics, what I'll call arguments from collapse, model confusion, and circularity. Section three is going to turn the tables by debunking a familiar consideration in favor of compatibilism rather, than incompatiblism which sets the stage for a fourth, I think harder, argument for incompatiblism according to which the best explanation of the various successes that mainstream semantics has enjoyed is a correspondence conception of truth, which is supposed to incompatible with deflationism about truth. And then in section four, which is really the heart of the paper, I'm going to argue that this inference to the best explanation argument fails to rule out two kinds of deflationism. In particular, a kind of deflationism tries to define truth in terms of deflated concept of referenc, which, Hartry Field famously accused Tarski of having, and a second kind of deflationism that I prefer which tries to invert the traditional order of explanation and give a theory reference taking the notion of truth for granted. Alright, so that's very abstract.

Section one, the fixed idea. So let me just pin down what some of the concepts I've been using already mean. To begin with, you might say deflationism about truth has two main strands, semantic and metaphysical, respectively. So the primary, semantic strands begins with some account of the meaning of “true” in cognate expressions, which typically emphasizes the intimate connection, as I said before, between affirming some arbitrary thing and affirming the thought that it's true. So typical deflationist sloganeering is something like our concept of truth is exhausted by the equivalence between saying “snow is white” is true, for example, and just saying “snow is white.” Applying the truth predicate in the sentence of our language as a way of indicating that it's a sentence we're prepared to assert or accept. So witness, for example, Quine's classic view that the truth predicate is merely a device for semantic ascents and may be enabling us to formulate generalizations and make blind endorsements that would otherwise have been beyond our linguistic abilities. So try saying, “Everything the Pope says is true,” without using the word “true,” for example. How else are you going to endorse everything the Pope says? Differences in detail between competing versions of deflationism are overshadowed by a common commitment to the idea that the meaning of “true” is exceedingly simple.

So the secondary, metaphysical strand of deflationism, is roughly that once we've explained what the word true means, there's little more to say about what truth itself is. So, unlike being a sample of water or having commercial value, being true is not a substantive property susceptible to the naturalistic reduction in more basic terms, or real as opposed to nominal definition. Indeed, some deflationists go so far as to say that there's no such property as being true, there's just the word. Anyway. Deflationism is therefore incompatible with paradigmatic versions of the so-called correspondence theory of truth, according to which being true is the property enjoyed by a sentence, or proposition, or belief, or other truth bearer when it can be decomposed into parts that pick out things in the world that are structured in a way that somehow mirrors the structure of the sentence itself. So this is sometimes called correspondence by congruence theory of truth, correspondence by structural isomorphism associated with many heavy weights in the history of analytic philosophy. There are other versions of the correspondence theory of truth besides the one that I reticulated, but only the congruence version, or the version of structural isomorphism is going to be relevant to what happens later in the talk. So incompatibilists, in the sense that I was using that term earlier, maintain again that deflationism about truth cannot be comfortably or coherently combined with the work-a-day business of mainstream semantic theorism.

So I've described mainstream semantics as truth conditional, model theoretic, formal, and linguistic. So it's all of those things, but those things are not synonyms for each other. So let me just clarify what my target is in the paper. So there are many different formal approaches to semantics, only some of which are truth conditional. And truth conditional semantics itself can be axiomatic as with Donald Davidson, or it can be model-theoretic as with Montague. And, of course, plenty of philosophers practice mainstream linguistic semantics part time, or more. So let me be clear that by mainstream semantics, I mean to pick out the leading approach to the theory of meaning and linguistics departments today, namely model theoretic semantics. I'll occasionally use the other words just so I don't keep repeating the phrase model theoretic semantics, but that's just loose talk.

So what is model theoretic semantics? It is a branch of applied set theory. So what the model theoretic semanticist does is associate declarative sentences with functions, in the mathematical sense, a function, from possible situations or worlds to truth values in an effort to capture their intuitive truth conditions and implication relations. This is a bit of an over simplification, but anyway. So these assignments of functions from possible situations to truth values to sentences, whose meanings you want to model or explain, are then considered in the light of deliverances from syntactic theory to generate hypotheses about the meanings or semantic values of significant subsentential expressions, or words appearing within the sentences. These, too, are construed as functions with variations in the domains and ranges dictated largely by the syntactic types of the expression in question, in the hope that the compositionality of meaning, the intuition that the meanings of big things are built up from the meanings of little things that make them up, can be secured by straightforward functional composition in a mathematical sense, a function composition. So truth is effectively rendered as truth in the intended interpretation and implication is understood in the standard Tarskian way.

One more preliminary, it's important to appreciate at the outset that the correspondence theory of truth is not literally part of model theoretic semantics so that the foregoing presentations of both model theoretic semantics and the correspondence theory of truth, may have sounded somewhat similar. Correspondence theory is an account of the nature of truth, whereas mainstream semantics is an account of the meanings of sentences and their parts. The latter uses the notion of truth and may, in some sense, as we'll explore, presuppose that truth itself can be analyzed along correspondence theoretic lines, but the science of semantics does not logically entail any particular metaphysics of truth. So if you don't want to take my word for it yet we'll come back to it in section four.

Section two. So these are the three arguments for incompatibility that I wanted to discuss and get out of the way quickly from collapse, modal confusion, and circularity. So a natural thought to have when first confronted with the foregoing descriptions of mainstream semantics and deflationism is that the former just collapses, that is semantics just collapses into the vacuous project of explaining tautologies when the latter, deflationism, is assumed. After all, the model theoretic semanticist aims to recover our intuitive judgments about truth conditions by appealing to theoretical hypotheses about the meanings of flexible units. The theory is adequate to the data, when it entails claims of the form “S is true if and only if P,” that's the semantic theorem scheme, where “S” is replaced by a term in the meta-language, a language which you're giving the semantics that denotes a sentence in the object language, the language you're interested in giving the semantics for, the language you're studying. And “P” is replaced by a sentence in the meta-language that intuitively has the same truth conditions as the sentence mentioned on the left-hand side. So if the truth conditional and model theoretic semanticists were studying French and giving his theory in English, an example -- an illustration of the scheme might be [speaking in French] is true if and only if snow is white. Something like that. But the argument from collapse, this is the first of the three arguments, continues. If deflationism is right then the left-hand side of any such biconditional will mean exactly the same thing as the right-hand side, and so the biconditional will be equivalent to the tautology “P if and only if P.” But mainstream semantics can't be in the business of explaining why tautologies hold so there must be something wrong with deflationists. Okay. I went through that very quickly because obviously it’s a terrible argument. I don't think anyone has ever been seduced by this argument. But I think focusing, getting clear on what exactly goes wrong with the argument will not only help pin down the content of deflationism, but also help set up the following two more interesting arguments for incompatibility.

So not since Ramsey have deflationists maintained that the two sides of such bi-conditionals, bi-conditionals in the semantic theorem scheme, have the same meaning. And even Ramsey would have balked at the suggestion in cases where the meta-language does not contain the object language, as when the English semanticist studies French. It can be conceded, of course, that when the meta-language does contain the object language, so when the English semanticist studies English, and when “S” is replaced by a quotation name for the very sentence that replaces “P” – so in effect when we're operating with this slightly different scheme. The contemporary deflationist will typically assert that the resulting instance of the scheme is in some sense analytic or true simply in virtue of the meaning of true. But notice this doesn't entail that the two sides of the biconditional are synonymous. It's very -- well, it's relatively, if you believe in analyticity it's very easy to generate analytic biconditionals. It can take two analytic truths, intuitively not to have anything to do with each other, and just put them on either side of a biconditional and you'll thereby have produced an analytic biconditional. You don't have to claim that the two sides of the biconditional are synonymous. But whether or not the deflationist can succeed in making his point, it's clear that deflationists shouldn't want to say that the left-hand side and the right-hand side of this biconditional, or for that matter the left-hand side and the right-hand side of this biconditional, literally mean the same thing, because the left-hand side is always about language. It has the quotation marks. Left-hand side's always about language. It's about the truth of a sentence. And the right-hand side is often not. In such cases, the right-hand side could have been true even if language had never developed. Okay. That's the first argument from collapse dispensed with.

This response to the collapse argument immediately raises what Claire Horisk calls the argument for modal confusion. It's occasionally been suggested, apparently by my dissertation advisor, Scott Soames, in his 1984 article, “What is a Theory of Truth.” It's occasionally been suggested that deflationism is incompatible with truth conditional semantics because of two theories attribute different modal profiles to instances of the “T” scheme. For the semanticist, “T” sentences, instances of the scheme, like quotes know as what is true if and only if “snow is white” express contingent facts about the meanings of sentences. Individually, orthographically, or phonetically, the sentence “snow is white” could have been used to mean something than what -- something other than what it actually means. But as we've seen, the deflationist typically thinks that many instances of the “T” scheme are analytic, true in virtue of the meaning of true, and analyticity is usually thought to entail necessity. So the deflationist has to say, it seems, according to the model confusion argument, that instances of this scheme, the “T” scheme, are necessarily true whereas, the semanticist doesn't want to say that. Semanticist is reporting contingent facts about the meanings of some language or other.

Okay, response. If the deflationist is being careful, however, she will not say that such a biconditional is true merely or wholly in virtue of the meaning of the word “true.” Another part of the reason it's true is that the sentence mentioned on the left-hand side is the same one used on the right-hand side. This is a semantic fact, too. It's a fact about denotation to be more specific. But it's a contingent semantic fact. Plus the deflationists can maintain that the relative “T” sentences are both analytic in a sense, the sense being true in virtue of semantic facts alone, and merely contingent. So if the quotation name in a given “T” sentence had referred to a different sentence, the “T” sentence might not have been true, but holding fixed the fact that it does refer to the sentence it actually refers to, the meaning of “true” does the rest of the work and ensures that the “T” sentence is true. Given all the relative semantic facts, not just those about “true,” the “T” sentence is guaranteed to be true, but of course those semantic facts could have been different. I should say this is a somewhat idiosyncratic conception of analyticity, but philosophers have learned to live with, more generally speaking, notions of analyticity that don't entail necessity. So, for example, we have a notion of Kaplan analyticity exhibited by a sentence like, “I am here now.” Simply any utterance of that sentence is guaranteed to be true just in virtue of how the [inaudible] expressions in that sentence work. But it's not a necessary truth that I am here now. I could not have shown up.

[Background laughing]

Okay, so more can be said for the argument for modal confusion and perhaps, even the argument from collapse, hopefully I've said enough at least to shift the burden of proof, but however that may be a further family of arguments that has a more urgent claim on our attention. Circularity arguments are more prevalent than any other style of argument for incompatiblism in the literature. So you see circularity arguments all over the place, in Davidson, in Dummett, in Gupta, in Collins. And despite our commitment to incompatiblism on other grounds, Horisk, whom I mentioned earlier, actually has argued persuasively that all extent circularity arguments fail. This is an excellent paper of Horisk from 2008 called -- called “Truth, Meaning, and Circularity in Phil Studies,” 2008. I highly recommend it. There's no real substitute for a careful reading of her paper, but you can convey a fairly accurate sense of what goes wrong with circularity arguments in fewer words. And I think it would be a tactical error to omit any treatment of circularity in a project like this one. The war of reconciliation must be fought on all fronts at once. Sorry, that's cheesy.

[Background laughing].

At a high level of abstraction all circularity arguments have the following form. The deflationist uses the notion of meaning to give a theory of truth, while the mainstream semanticist uses the notion of truth to give a theory of meaning. You can't have it both ways on pain of circularity. Either meaning is explanatorily prior to truth or truth is explanatorily prior to meaning. Now interestingly enough, versions of this argument have been endorsed by deflationists and their opponents alike. So Paul Horwich, famous deflationist, formulates the point thusly, “Trying to squeeze both deflationism and truth conditional semantics out of the ‘T’ scheme is like trying to solve an equation for two unknowns.”

[Background laughing]

So different versions of the circularity argument emphasize different notions of priority, but we can afford to paint with a broad brush for present purposes. So let me grant that the deflationists will indeed have to use the notion of meaning, or some related semantic notion, in specifying the meaning of true. After all, the deflationist can't simply say that the notion of truth is implicitly defined by the sentential “T” scheme, he'll also have to tell us what sorts of things can be substituted for the dummy letters in the scheme. So presumably, imperatives and interrogatives are out of bounds, noun phrases and predicates, too.

Perhaps, these linguistic entities can be excluded on purely syntactic grounds, but what about jabberwocky sentences? Green ideas leap furiously. Or other indicatives the philosophers have been inclined to describe as failing to express proposition. So examples from ethical discourse might be pertinent here. If the deflationist wants to exclude these sentences as inapt for substitution in the “T” scheme, she'll be forced to use some semantic notion in her specification of the meaning of true. She's saying effectively the only thing you can sub in for “S” -- the only things we can sub in for “S” – are sentences that are properly meaningful. So grant that. Grant also that the model theoretic semanticist takes meaning to determine truth conditions. So for the model theoretic semanticist to say that the sentence, for example, “Green ideas leap furiously,” is not suitable for substitution in the “T” scheme because it's meaningless, is therefore to commit herself to the view that it lacks truth conditions. Thus the objection from circularity can be unpacked as the charge that the aspiring model theoretic deflationist, the person who's trying to put these two things together, would have to use the notion of truth, because she'd have to use the notion of truth condition, to specify the meaning of the word “true.” That sounds bad.

Notice, though, that deflationism is a complete red herring here. The argument really doesn't have anything to do with deflationism. The sort of circularity at issue will afflict any mainstream account of the meaning of the word “true,” or the meanings of sentences involving the word “true.” So this observation might motivate Davidsonian primitivism about truth, but that seems to me like an overreaction. The question is, is mainstream model theoretic semantics really constitutionally incapable of specifying the meanings of truth descriptions? No, it's not. Suppose is true functions is a predicate in natural language, which most deflationists and all, I think, inflationists alike will typically allow. And suppose for the sake of simplicity, but without loss of generality, that the model theoretic semanticist just takes the meanings of semantic values of predicates to be their extensions, the sets of things to which they apply. Thus we can say that the meaning of the truth predicate in any given language is just a set of all and only those sentences of that language to which the truth-conditional semanticist would assign the value one relative to the actual world. Not only the proposal informative, it seems completely in keeping with the spirit of deflationism. This is going by a bit quickly, so we might want to return to this paragraph.[Laughs] So I go on to say, there will be complications with this proposal if the language in question has the resources to express liar style sentences, like this sentence is not true. But that's not a special problem for a model theoretic formulation of deflationism about truth. In fact, if we try to treat the paradoxes with a Tarskian or Kryptian [phonetic] hierarchy of languages, circularity is going to be avoided altogether by distinguishing the object language truth predicate from the metalanguage truth predicate.

One might object. A guy called Doug Patterson has, in fact, objected (in 2005) that the metalanguage truth predicate is not deflationary, but it seems to me there's no obvious reason why we simply -- we can't simply iterate the modal theoretic version of deflationism from the previous paragraph to give the meaning of the metalanguage truth predicate in a meta metalanguage and so on.

[Background laughing].

Yeah, it's weird. So in the Patterson paper he grants that the truth-conditional semanticist can be a deflationist about object language truth predicates, so object language meaning the language that you're studying, but insists she just can't be a deflationist about the truth predicate used in her semantics. And then there’s – the paper – it goes on for many pages but there's no argument. There's a gap where the argument should have been. This is prejudicial perhaps, but I think that paper is worth thinking harder about. Okay. That's end of section two. So that was my reaction to, sorry guys, I know, that was my reaction to the first three arguments for incompatibility. Went by pretty quickly.

Now what I'm going to do in section three, just to remind you, is try to motivate a fourth argument for incompatibility, but it's going to start by rehearsing an easy argument for compatibility between deflationism and semantics that I think doesn't work. And seeing why it doesn't work I think helps motivate the real argument for incompatiblism. Okay, that was a little bit convoluted. Anyway, I'll just get into it. While I do think circularity arguments fail to establish incompatiblism, I don't think that compatiblism is easy to establish.

So it's been often noted that nothing hangs on how we label the entities in the range of the functions that model theoretic semantics associates with meaningful declarative sentences. We can call them true and false, we can call them one and zero, we can call them chocolate and vanilla, it doesn't seem to matter. So the appearance that model theoretic semantics, according to this argument, has something essentially to do with truth, it's just a typographical illusion perpetuated by the same mechanisms that maintain all linguistic conventions. So it would be therefore, hard to see how mainstream semantics could be incompatible with deflationism about truth. The two theories would simply concern different topics. But the problem with this argument for compatiblism is that zero and one, or whatever terms we use, are supposed to denote things that play a central role in explaining why we tend to accept the sentences that we do in irrelevant circumstances and why we tend to infer the conclusions that we do from the relevant premises. The guiding presumption of mainstream semantics is that these patterns of inference and acceptance somehow arise from our ability to recognize implications and truths in context. If that presupposition is justified, then it would seem that the entities in a range of the functions that model theoretic semantics associates the meaningful declarative sentences have got to be truth in falsity or at least properties necessarily coextensive with them. Granted, we can call truth whatever we like, but a rose by any other name would not only smell just a sweet, but in fact still be a rose. So naturally, many philosophers of language would question what I've been calling the guiding presumption of mainstream semantics.

Why not think instead, that the relevant data results from our ability to recognize when acceptance and inference would be dialectically appropriate, or epistemically justified as something else altogether? On the other hand, if the presupposition is right; if for example, our inference patterns really are to be explained by appeal to some knack we have for recognizing implication relations when they obtain, then presumably we should look to logic and epistemology to explain those data, not to the theory of meaning. At best, only the semantics of illogical particles would be involved in the explanation. So objections like these are often used to motivate, so-called use theoretic alternatives to mainstream semantics, departing radically from the orthodoxy and explanatory ambitions and means. Notice, however, that such objections could conceivably be used instead, to motivate a much more modest use theoretic cooption of the formal apparatus of mainstream semantics. There's no obvious reason the use theorist can't formulate her position in the language of set theory. So suppose, for example, that a use theorist thought, like Michael Dummett, but unlike Paul Horwich, that it makes perfect sense to talk about complete sentences being used correctly or incorrectly. So certain circumstances in which it's appropriate to say that “the cat is on the mat” and certain situations in which, it's inappropriate to say “the cat is on the mat.”

Truth doesn't enter into it, according to the use theorists. Then the use theorist can simply take the values and the range of the functions that the standard view associates with sentences to be assertability and unassertability rather than truth and falsity. So crucially, this suggestion is not merely terminological as before; it concerns what the properties in question are. So the truth condition semanticist assigns sets of possible worlds or situations to sentences and, or pairs of sets, and labels one “T” and the other “F.” We can use the apparatus of set theory and still talk about sets of possible situations, but not use the notion of truth if we’ve got different background presumptions about how meanings can be explained. In particular, we can think of – we can model – the meaning or semantic properties of sentences as sets of worlds understood as assertability conditions, or acceptance conditions, and unassertability or unacceptance conditions.

But here's the problem, one still wants a systematic explanation of why any given sentence means what it seems to as a matter of theoretic intuition. The standard explanation evokes the compositionality of meaning. Meanings of big things derive from the meanings of their parts. Which model theory, and this is why it's been so useful, can be used to formalize. But in order for the use theorists to coop this formalism, she'll have to give a use theoretic account of the sorts of functions that orthodox semantics associates with significant subsentential expressions and it's just entire -- and turns out to be entirely unclear how to do that. The situation is so bad that the most prominent use theorist, Paul Horwich, has, as you may know, produced a series of papers in which he's argued that the principle of compositionality imposes no substantive constraint on the theory of meaning, which to people like Jason Stanley seems absurd. This is a problem. So herein lies the chief advantage of mainstream semantics. You've got a conception of what the meanings of sentences are, and you've got a conception about what the meanings of parts of sentences are, and those two conceptions go hand in hand through the formalism that is the apparatus of model theoretic semantics. Okay.

So here's the best argument for incompatiblism I can think of. If the meanings of significant subsentential expressions could be identified with functions ultimately built up from objects, properties, events, processes, and other materials out there in the world, and if the meanings of sentences could be identified with their truth conditions then some version of the correspondence theory of truth could do double duty as a guide to the mechanics of semantic composition. After all, the correspondence theory purports to tell us how the truth conditions of sentences depend upon the broadly referential relations that hold between their constituent words and entities out there in the world. If the correspondence theory were right, then the model theoretic semanticist would have a familiar recipe for composition already in hand. Indeed, the fact that model theoretic semantics has enjoyed so much success could be explained in part by the fact that truth amounts to correspondence to reality rather, than what the deflationist wants to say it does.

Okay, so this last sentence encapsulates what I've been calling the inference, what I called earlier, the inference to the best explanation argument for incompatiblism. The explananda, the things to be explained, are the various successes of mainstream semantics, which somewhat confusingly are themselves explanations of our intuitions about truth and implication. And the explananda, the thing doing the explaining, is the correspondence theory of truth. If this instance of inference to the best explanation is good I think there's a tangible sense in which mainstream semantics presupposes the correspondence theory of truth and therefore, tangible sense in which deflationism is incompatible with semantic orthodoxy, despite the failure of the three arguments that we discussed earlier. So the burden of the next section, the main section of the paper, is going to be to show that the inference does not, in fact, go through. Skip some of this.

Okay, let me just say what the main argument again is, that is the main argument that I want to reject or debunk. It seems to me that the best reason for thinking mainstream semantics can't plausibly be combined with deflationism is that the rival correspondence theory of truth provides a better explanation of the striking successes for the model theoretic orthodoxy. Compositionality of model theoretic meaning mirrors beautifully the correspondence theorist's definition of truth, in terms of reference, application, and related notions. Alright, response.

So there are at least two recognizably deflationary strategies for resisting this inference to the best explanation. The first involves arguing that the theory of truth Field famously accused Tarski of having, explains the successes of model theoretic semantics just as well as any correspondence theory. Most contemporary deflationists will be reluctant to embrace this strategy as it involves conceding that there is much more to our ordinary notion of truth, than a “T” scheme itself.

The second option, the one that I prefer, involves inverting the correspondence theorist's order of explanation, effectively defining reference in terms of a deflated concept of truth. As I said, this is the strategy I ultimately endorse and recommend even to those deflationists about truth who would like to be deflationists about reference as well. Alright, so let me make good on the promissory notes made in that last paragraph. So Hartry Field famously argued that Tarski failed to satisfy his own reductive physicalist ambitions in giving a theory of truth, by ignoring the need to provide a substantive theory of denotation to compliment his semantic conception. What Tarski does say about denotation looks a lot like what contemporary deflationists about reference said, that the concept of reference is in some sense exhausted by a schematic platitude like “T refers to T.” Quotes on the left-hand side, no quotes on the right-hand side. So this is an interesting note about Field's psychology, I haven't asked him about this, I probably should. Field at this point in his career saw this as a defect in Tarski, because he was still trying to be a correspondence theorist, Field. He was looking to Tarski for help. But later in his career Field came around finally to being a deflationist about truth, rejecting correspondence theory, but it seems not to have, maybe it did. I shouldn't say that, I'm sure it occurred to him, Field is incredibly smart, I'm sure it occurred to him to consider this other form of deflationism, a kind of deflationism that defines truth in terms of reference and related notions, but then goes on to give a deflated account of the nature of reference, something disquotational like “T refers to T.” That is not the kind of deflationism Field ultimately went in for, and I'm not quite sure why.

Anyway, whether or not Field was right about Tarski he provided the aspiring deflationist with a way to ape the correspondence theorists’ definition of truth in terms of reference. Elsewhere, I've argued that this admittedly unusual style of deflationism can be independently motivated by reflection of the problem of truth in value gaps, which I won't get into. But the pertinent question at present is – or the pertinent questions at present are – does the view of Field-Tarski, provide any explanation of the successes of mainstream semantics? And if so, how does it compare in this respect to the correspondence theory of truth?

So one might argue that a deflationism about truth, defines truth in terms of the deflated conception of reference, does not provide an explanation of the successes of model theoretic semantics for the following reasons. Well, you might argue that mainstream semantic presupposes some inflationary account of reference, some semantic values or set theoretic entities built up from objects, properties, and other entities out there in the world. So presumably we won't need some substantive scientifically respectible theory of word-world relationships if we're to explain why linguists have had so much success accounting for truth conditions and implication relations compositionally. Whether or not we consider inflation as a matter of reference to be part of the correspondence conception of truth, which may ultimately be a terminological matter, deflationism about reference is an essential part of the version of deflationism about truth currently under consideration. So if an inference to the best explanation argument supports inflationism about reference, then the deflationism of Field's-Tarski should be rejected. But does an inference to the best explanation argument really support inflationism about reference? After all, so here's the crucial point, the deflationists and the inflationists about reference, so the deflationist is the guy, the person who thinks that the notion of reference is exhausted in some sense by something as seemingly trifling as the schematic claim that “T refers to T,” “Sherlock Holmes refers to Sherlock Holmes.” A problem there already, right? Sherlock Holmes doesn't exist. Modifications can be made.

[Background laughing].

The inflationist is the person who thinks that no, we need a science project to figure out what this thing – reference – is like. How do we explain the fact that creatures like us using language or thought can represent the world, how we look at a sentence? “Snow is white” is true if and only if the predicate is “white” applies to the reference of “snow.” You see there are variations of sentences like this which are, typical of mainstream semantic theorizing in cartoon form in deflationary tracks. And the derivations proceed by using instances of the “T” scheme and instances of the disquotational scheme for singular terms and general terms. At the very least, present day deflationists should grant that material by conditionals like this one are true, even if not analytic of truth, even if they don't get at the meaning of the word “true.” So this observation might inspire a different strategy for reconciling deflationism about truth and mainstream semantics to argue that inference to the best explanation merely establishes biconditionals linking truth and reference, not to further claim that truth can be, truth itself, can be defined along the lines suggested by the biconditionals. So however exactly the details are worked out, incompatibilists might object to this strategy on the following grounds. And this is the part that I'm most interested in.

So mainstream semantics takes word-world relationships to be somehow more metaphysically basic or fundamental than, the truth conditions of complete sentences. And the correspondence theory of truth seems to vindicate that outlook as it, too, takes broadly referential relations to be more basic or fundamental than truth, falsity, and other analytic phenomena. To put it metaphorically, the successes of model theoretic semantics lend indirect support to the correspondence theory because their arrows of comparative fundamentality point in the same direction. This is a metaphysical point. Field's-Tarski at least tries to imitate this order of explanation, but more popular versions of deflationism just leave no room for it. Derivations like the one on the board, or the one alluded to with effect of the sentence on the board, merely accentuate the fact that contemporary deflationism doesn't count in any essential conceptual or metaphysical connections between truth and reference. The derivation proceeds by independent assumptions about truth and reference.

Okay. Now there are countless ways in which one thing can be more or less fundamental or basic than another. In fact, we've already encountered two different kinds of comparative fundamentality within the practice of mainstream semantics itself, at the very start. On the one hand, the truth values of complete sentences are usually treated as being more epistemically, or evidentially fundamental than the meanings of their parts, insofar as linguists tend to use intuitions about truth values that justify hypotheses about the meanings of subsentential expressions and not the other way around. Though that does happen, too, sometimes. At the same time, the truth values of complete sentences are usually treated as being less theoretically or explanatorily fundamental in the meanings of their parts, insofar as linguists tend to answer why questions about the former with claims about the latter, rather than vice versa. And, of course, there are also psychological or cognitive notions of fundamentality at play in contemporary linguistics which have to do with the order in which meanings tend to be learned, and the order in which they're re-grasped by individual occasions. So of these various notions of fundamentality native to mainstream semantics the one relevant to the foregoing argument for incompatiblism is what I call theoretical or explanatory fundamentality.

So the question – Is this the same kind of comparative fundamentality at play in the correspondence theory of truth? No, I don't think so, for something like the reasons I'm about to utter, but not exactly. I've had an opportunity to rethink this part of the paper and maybe in Q&A it'll come out how exactly I'm thinking about things now. But anyway, let me say what's on the page. No. [Laugh]

So theoretical or explanatory fundamentality as we can understand, with respect to mainstream semantics, is just not the same notion as metaphysical, the notion of metaphysical comparative fundamentality native to the correspondence theory of truth. So semantics is a branch of applied mathematics, a flourishing scientific enterprise that uses formal tools to generate theoretical explanations of various phenomena having something or other to do with linguistic meaning. The correspondence theory of truth by contrast is a piece of analytic philosophy, a metaphysical theory that reports to explain what it is, for sentences in other truth theories to have a given truth value. Scientific explanation, or at the various different -- many different species of scientific explanation and metaphysical explanation are at least notionally distinct, explanations or answers to why questions usually, but there are many why questions that only scientists address and many why questions that only metaphysicians address. And this holds true even when we focus on topics of interest both to scientists and metaphysicians, like the nature of time, the individuation of biological species, the relationship between mind and the body, or the mechanisms of free choice. Of course, naturalist methodology requires that our answers to metaphysical questions on such topics comport with current scientific wisdom, but mere compatibility with science leaves a lot of room for heady metaphysical theory. All of this is just to say that metaphysical explanation is not the same concept as scientific explanation, and this is the sort of blunt claim that I want to qualify in the Q&A if I can. [Laugh] Really, I think the claim that I should be making is that constitutive explanation, which can come up both in science and in metaphysics, constitutive explanation is distinct from unification constitutive of explanation, and I think what's really going on in formal semantics is unification, but we'll say more about that later.

Therefore, the concept of explanatory priority at play and the correspondence theory of truth cannot simply be identified with the concept of explanatory priority, an issue in mainstream semantics. But this is just a conceptual point. So the conceptual point raises the further more substantive question, what bearing does the direction of the arrow of fundamentality and the metaphysics of truth have on the direction of the arrow of fundamentality in a science like semantics? Is it a good thing for them to point in the same direction or not? That's the question.

So in order to address this question we might compare the correspondence theory of truth with accounts of the natures of truth in reference that invert the correspondence theories metaphysical order of explanation. So Michael Williams in a great paper, “Meaning and Deflationary Truth,” '99, Journal of Philosophy, reads Donald Davidson as defending precisely such an inversion. So here's a quote from Williams on Davidson: "Reference relations for Davidson are not independently existing objects that explain observable features of a person's linguistic behavior, rather," and this is the crucial bit, "reference relations are constituted by their role in making a person's utterances, or most of them, come out true." This is a metaphysical story that says reference is derivative from truth, not vice versa, in the way that the correspondence theorists would have us think. The view articulated here in this quotation, whether or not it really belongs to Davidson, doesn't reject the methodology of contemporary semantics. After all, Davidson is, of course, one of the founders of truth conditional tradition in contemporary semantics. The point just rejects reading on metaphysics of truth off of that scientific methodology. So even if referential relations are scientifically or unificationally more fundamental than truth conditions, if in this respect they really are quite like electrons,they may still be less –be metaphysically less – fundamental in that they're somehow constituted by truth conditions.

So this is the main idea of the present paper. It's already in William's, but it's not been adequately appreciated in subsequent literature. I think this is important because the point, as he presents it, can be mistaken for just a mere scholarly observation about Davidson's philosophical system. On the contrary, the general moral is that inference to the best explanation arguments for incompatiblism just failed to rule out versions of deflationism about truth, that simply invert the correspondence theorist's order of explanation. So granted, standard biconditionals linking truth and reference, or reference-like notions, can look a lot more like definitions of truth than definitions of reference. But as we know, they can easily be converted into explicit definitions of reference and related notions using a variant on the Ramsey Lewis method for defining theoretical terms. So I say a variant just because the idea would be to replace the notion of reference, but not the notion of truth with a second order variable and then prefix the statement with just a single definite description forming operator.

So conceptual or inferential semantics provides a more familiar instance of this kind of inversion – truth first, reference second; crowding facts about reference and the truth conditions of and implication relations between complete declarative sentences, but this kind of role theory is usually advanced with bells and whistles that obscure its underlying compatibility with mainstream model theoretic semantics.

So the punch line is just this, if we can distinguish metaphysical from scientific explanatory priority or if we can distinguish constitutive explanation for unification explanation and make the claim that the kind of explanation at play in contemporary mainstream semantics is unification explanation, then we can reconcile standard versions of deflationism about truth. So forget about Field's-Tarski, nobody likes that. Standard versions of deflationism about truth with the practice of mainstream semantics. If biconditionals like “S” can be read as revealing something about the nature or essence of reference rather than the nature of truth, then there should be no semantic obstacle to embracing some standard deflationary theory of truth.

One worry you might have about this program is well haven't we just moved the bulge in the carpet? Aren't we inflating the notion of reference? I've got another paper in which I try to give you a theory of reference in terms of truth that looks a lot like standard deflationism, but is interesting and different in certain ways, but I'm not going to rehearse that other work here.

I think that's all I wanted to say. I've got a polemic at the end of the paper against the -- despite the risk of undercutting the significance of the present paper, I feel obliged to admit that I'm not convinced mainstream model theoretic semantics provides the best framework for theory of meaning, but of course that's beside the point. The point of the talk has been to say that [background laughing] there's no conflict between deflationism, the philosophical tradition, deflationism about truth and perhaps reference on the one hand, and the practice of work-a-day formal semantics on the other. So thanks for your attention. [Clapping]

==== Transcribed by Automatic Sync Technologies ====