

*Metasemantics and Singular Reference**

ORI SIMCHEN

University of British Columbia

1. Introduction

Metasemantics is concerned with how the semantically significant becomes endowed with its semantic significance. If semantics is of the usual truth-conditional sort, then a principal issue for metasemantics is how subsentential expressions become endowed with their distinctive contributions to the truth-conditions of whole sentences in which they partake. Metasemantics as it is ordinarily understood is the metaphysics of semantic endowment, where the latter is cast in truth-conditional terms.¹ It is beholden to semantics insofar as it attempts to articulate determinants for semantic endowment.

There are distinct ways of conceiving of the metasemantic project, however. An important choice point concerns whether to think of semantic endowment as emerging directly from conditions surrounding the production or employment of the items semantically endowed (e.g. causal relations borne to portions of speakers' environment), or to think it in terms of conditions surrounding the interpretive consumption or reception of such items (e.g. facilitation of good explanations of speakers' verbal behavior). The first general approach—call it *productivism*—is taken by the likes of Donnellan, Kaplan, Kripke, and early Putnam, among many others. The second approach, *interpretationism*, is the one usually associated with Davidson and Lewis. My aim is to offer a partial articulation and defense of a general productivist orientation to the subject by arguing against metasemantic interpretationism. I will make a case for the claim that an interpretationist orientation to metasemantics is severely flawed when it comes to singular reference. My argument will be cast against a Lewisian version of metasemantic interpretationism commonly known as reference magnetism, but the considerations I offer are general, so in the Appendix I adjust my overall argument to suit a Davidsonian framework as well. I end by drawing a general moral for metasemantics and its relation to truth-conditional semantics.

From a productivist standpoint, metasemantics primarily targets conditions of producing or employing an item of significance. In general, the sort of production metasemantics is concerned with is production of items of significance *qua* significant. In a different terminology, we are concerned with production of symbols rather than that of signs.² What distinguishes productivism as a metasemantic orientation is that the item's production *qua* significant depends directly on conditions surrounding the item's production or manipulation by the speaker or writer *qua*

insignificant. What makes it the case that a spoken referential token of a noun, say, has the significance it has, is determined directly by the circumstances under which the token was phonetically produced. Some productivist approaches appeal to referential intentions of speakers to explain how a referring token employed on a particular occasion comes to refer to what it refers to, as in Donnellan's (1966) treatment of referentially used descriptions or Kaplan's (1989) treatment of demonstrative pronouns.³ Other productivist approaches appeal to the causal history of the item of significance vis-à-vis the thing to which it refers without particular emphasis on referential intentions. So for certain versions of the approach what determines the semantic contribution of a token to truth-conditions on a given occasion of use is the referential intention with which it is employed or something similar, whereas for other versions what determines it is a more basic causal dependence of the representation—be it a token of a linguistic type or a particular mental representation—on the item(s) represented.⁴

Productivism may be contrasted with interpretationism, a metasemantic orientation whereby endowment with semantic significance emerges directly from conditions surrounding the interpretive consumption of the items thus endowed. Under this general rubric we have in the first instance the Davidsonian appeal to considerations of fit of assignment of semantic values to subsentential expressions with speakers' overall linguistic behavior and attitudes in context—the latter themselves subject to further interpretability—so as to engender explanatorily suitable ('interpretive') truth-conditions for whole sentences in context. But my focus here is on the Lewisian appeal to a constraint of worldly naturalness in the assignment of semantic values to subsentential expressions that maximizes overall truth for the global theory in which the expressions embed. Beyond important differences among interpretationist approaches, an underlying theme that unites them is that endowment with significance is a matter of the items thus endowed being regardable in a certain way, whether such regarding facilitates an account of the rationality of the speaker's linguistic behavior in his or her worldly surroundings (Davidson), or whether it also facilitates an explanatorily superior theoretical capture of the way the world is (Lewis). Either way, endowment with significance does not emerge directly from conditions surrounding the production or employment of the item in question but from conditions surrounding their post-production assessment.⁵ In Lewis's case the shape of the world and the shape of our theory of the world together conspire to make it the case that our predicates have the significance they have—a post-production affair.

In what follows I will argue that the latter approach comes up short in handling a certain under-explored threat of referential indeterminacy, a kind of indeterminacy that targets singular reference in particular. The bottom line will be that (1) referential indeterminacy of the type to be discussed arises only in an interpretationist setting—productivism is immune to it—and (2) a Lewisian antidote in terms of eligibility of interpretation characterized by appealing to naturalness for properties is incapable of dealing with the problem.

2. Reference Magnetism

I begin by sketching my primary target. Lewisian reference magnetism is in the first instance a thesis about the semantic values assigned to predicates.^{6,7} The idea, in a nutshell, is that certain overall interpretations of our language into the world are objectively better than others due to their higher eligibility, the latter understood in terms of maximization of naturalness in the assignment of semantic values to predicates. The thesis was initially proposed as an antidote to Putnam's model-theoretic argument, an argument purporting to show that under certain minimal assumptions about realist truth, the distinction enshrined by realists of all stripes between epistemic ideality and metaphysical truth cannot be sustained. Putnam's argument turns on the (almost) inevitable availability of an overall interpretation of our language into the world that renders an epistemically ideal theory true of the world. The argument exploits a basic point about model-theoretic interpretation.

Assume with the realist that the world is a totality of mind-independent things. (For present purposes we need not enter the fray of trying to precisify the dark notion of mind-independence.) Let T be our epistemically ideal theory in a first-order extensional language. T would be at least consistent, so it would have a model. Under certain minimal assumptions about T and the size of the world, T would have a model m of exactly the same size as the world. By exploiting the existence of a bijection from the domain of m into the world itself we can define a model m^w of T that has the world itself as its domain. So the epistemically ideal T turns out to be true of the world after all—there is no way for it not to be true of the world under minimal assumptions. The distinction between epistemic ideality and realist truth collapses.

Here is how Lewis (1984) responds to this argument in terms of eligibility: “When we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory” (227). Let us flesh this out a bit. The interpretation Putnam's argument appeals to in forcing the pronouncements of epistemic ideality to come out true of the world may very well assign, if we happen to be epistemically unlucky, highly gerrymandered semantic values to our predicates that do not respect objective joints in nature. Recall that m^w was defined in terms of a bijection from the domain of m into the domain of m^w . The bijection itself was arbitrary; all it did was ensure that the structure imposed by T on m is replicated in m^w regardless of independent features of the individuals in the domain of m^w . But the domain of m^w is just the totality of worldly things. From the Lewisian standpoint the misstep in Putnam's argument is the failure to distinguish arbitrary interpretations of T from intended ones, ones that respect the structure that already inheres in the world itself. It is only the latter that are relevant for the assessment of the realist point that T might be false of the world. T would be false of the world if it so happens that it has no model isomorphic to the way the world really is.⁸

This is undoubtedly a formally adequate response to Putnam's argument.⁹ It relies on a certain idea that those with Kantian leanings may find spooky and unilluminating—the idea that the world has its own inherent structure

independently of our conceptual involvement in it.¹⁰ The pros and cons of this idea lie beyond our present concern. The question before us is whether anything in this proposed antidote to Putnam's argument touches on reference understood in its common singular variety. Can Lewisian considerations effectively block indeterminacy in singular reference? Suppose I say 'This is a nice piece of fruit' with respect to a particular apple in a fruit stall, a paradigm case of singular reference. I would have thought that I was referring to the particular apple in the stall. Can Lewisian considerations rule out that I was actually referring to the exoplanet Alpha Centauri Bb, some four light years away, or to the number 17?¹¹

An initial thought is that they can. Consider the following passage from Lewis (1983):

Naturalness of properties makes for differences of eligibility not only among the properties themselves, but also among things. Compare Bruce with the cat-shaped chunk of miscellaneous and ever-changing matter that follows him around, always a few steps behind. The former is a highly eligible referent, the latter is not. . . . This is because Bruce, unlike the cat-shaped chunk, has a boundary well demarcated by differences in highly natural properties. Where Bruce ends, there the density of matter, the relative abundance of the chemical elements, . . . abruptly change. Not so for the chunk. (372)

There is a metasemantic idea implicit here that needs to be drawn out. How differences in eligibility for properties are supposed to carry over to differences in eligibility for things is illustrated by comparing Bruce the cat with a cat-shaped chunk of matter following Bruce around (assuming the chunk is not just another cat stalking Bruce). Natural properties mark Bruce's boundary, we are told, but the same cannot be said for the relatively arbitrary chunk behind Bruce. Suppose this is so. Starting with a corpus of sentences held to be true, let one be 'Bruce has an organic surface', where 'organic surface' applies to the outermost 10 micron thick layer of a more or less contiguous body mostly constituted by organic molecules. Assuming that 'organic surface' already stands for a relatively natural property, 'Bruce' will refer to the cat rather than the relatively arbitrary chunk of matter behind him on pain of falsifying the sentence. But this strategy of ruling out the chunk in favor of Bruce as the intended referent for 'Bruce' extends beyond cases where the choice is between a cat and a gerrymandered chunk of matter. It generalizes to cases where the choice is between two things whose respective boundaries are equally well demarcated by differences in natural properties. Consider Bruce and Bruce's stalker Lenny, a cat always three feet behind Bruce. Holding 'Bruce has a center of mass at x, y, z, t ' to be true and holding the interpretation of the predicate fixed will decide whether 'Bruce' refers to Bruce or to Lenny, depending on which of the two cats has a center of mass at x, y, z, t . So the more general metasemantic idea implicit in this talk of eligibility for things is that holding the interpretation of predicate letters fixed will also fix the interpretation of singular terms, given a certain allocation of truth-conditions to sentences.

Go back to the earlier example of 'This is a nice piece of fruit' where the intended referent for the relevant occurrence of 'this' is an apple rather than a planet or a number. Any candidate for being the intended interpretation of the language would

need to respect that. Consider three rival interpretations. Each assigns the set of all and only pieces of fruit to ‘is a piece of fruit’, the set of all and only planets to ‘is a planet’, and the set of all and only numbers to ‘is a number’.¹² But one interpretation assigns the apple in the stall to ‘this’ as used on that particular occasion. A second assigns Alpha Centauri Bb. The third assigns the number 17. Only on the first interpretation will ‘This is a piece of fruit’ be true if and only if the apple is a piece of fruit. On the second interpretation ‘This is a piece fruit’ will be true if and only if Alpha Centauri Bb is a piece of fruit. On the third interpretation ‘This is a piece of fruit’ will be true if and only if the number 17 is a piece of fruit. So on the assumption that allocation of truth-conditions to sentences is to be as of the first interpretation, we can rule out the second and third interpretations as unintended. Considering only these three rival interpretations, we seem to have managed to secure the determinacy of ‘this’ as used on the relevant occasion to refer to the apple rather than to Alpha Centauri Bb or the number 17. So far, so good.

Here, however, is a nagging thought. Consider the second interpretation, the one assigning Alpha Centauri Bb to ‘this’ as used on that particular occasion. We assumed that on the second interpretation ‘This is a piece of fruit’ comes out true if and only if Alpha Centauri Bb is a piece of fruit, ‘This is a planet’ comes out true if and only if Alpha Centauri Bb is a planet, and ‘This is a number’ comes out true if and only if Alpha Centauri Bb is a number. This is because we assumed that the way in which ‘This is a piece of fruit’, for example, comes out true on the first interpretation is the same as the way in which ‘This is a planet’ comes out true on the second interpretation; and we assumed that the way in which ‘This is a planet’ comes out false on the first interpretation is the same as the way in which ‘This is a piece of fruit’ comes out false on the second interpretation, and similarly for the other cases. In short, we assumed that being true or false for whole sentences, as a function of the interpretation of subsentential expressions, is univocally fixed. But this might be challenged in turn. What if truth for sentences as a function of the interpretation of subsentential expressions is itself open to indeterminacy? The suggestion may seem outrageous, sure, but no more so than the original suggestion of referential indeterminacy. If the latter got a philosophical hearing, then the former should get one as well. Might it be that for a sentence of the form ϕt to be true is for something *other than* the intended interpretation of t to be among the things in the intended interpretation of ϕ ? If so, then the strategy appealed to above for ruling out the second and third interpretations as unintended may not be available after all. For suppose that it is a feature of truth for sentences—as opposed to a feature of the interpretations of ‘this’ and of ‘is a piece of fruit’—that ‘This is a piece of fruit’ comes out *true* on the interpretation that assigns Alpha Centauri Bb to ‘this’ and assigns all and only pieces of fruit to ‘is a piece of fruit’. And suppose, correlatively, that it is a feature of falsity that ‘This is a planet’ comes out *false* on the interpretation assigning Alpha Centauri Bb to ‘this’ and all and only planets to ‘is a planet’. If these are somehow live options, then they spell trouble for the above strategy of ruling out the second and third interpretations as unintended due to getting the truth-conditions wrong.

3. Scrambled-Truth-in-a-Model

Before discussing these options further let us fill in some of the details by focusing on a simple first-order extensional language L . Besides the usual first-order logical vocabulary, L contains only constants c_i and predicate letters of various arities P_j^n . A model m is defined in the usual way as $\langle M, \mathfrak{I} \rangle$, where M is a universe of discourse and \mathfrak{I} an interpretation function that assigns to each constant t a member of the universe M and to each predicate letter ϕ of arity n a subset of the n^{th} Cartesian power of M (an n -place relation).

We have the familiar inductive definition of *truth-in-a-model* (\models). Let us only concern ourselves with the atomic sentences $\phi(t_1, \dots, t_n)$. The relevant clause in the definition of \models , where m is $\langle M, \mathfrak{I} \rangle$, is:

$$m \models \phi(t_1, \dots, t_n) \text{ iff } \langle \mathfrak{I}(t_1), \dots, \mathfrak{I}(t_n) \rangle \in \mathfrak{I}(\phi).$$

In the full definition of truth-in-a-model a provision would be made for the assignment of values to free variables before a general clause for atomic formulas is given, a point to which I return briefly below.

Now let m^L be $\langle M^L, \mathfrak{I}^L \rangle$, where M^L is the intended domain ('L' for 'Lewis'). We assume that for any \mathfrak{I} , $\mathfrak{I} \neq \mathfrak{I}^L$, \mathfrak{I} is no more eligible than \mathfrak{I}^L as an overall interpretation of the language when it comes to the predicates. \mathfrak{I}^L is thus maximally eligible by Lewisian standards.

Now a preliminary observation: there is an interpretation \mathfrak{I}' , $\mathfrak{I}' \neq \mathfrak{I}^L$, that agrees with \mathfrak{I}^L on the assignment to every predicate letter in the language. Simply define $\mathfrak{I}'(\phi) = \mathfrak{I}^L(\phi)$ for every predicate letter ϕ and let $f: M^L \rightarrow M^L$ be a nontrivial permutation on M^L , defining $\mathfrak{I}'(t) = f(\mathfrak{I}^L(t))$ for every constant t . Where m' is $\langle M^L, \mathfrak{I}' \rangle$, there is no guarantee of course that the same sentences will come out true in m^L and m' .

We now define a new notion: *scrambled-truth-in-a-model* (\models^σ). For $m = \langle M, \mathfrak{I} \rangle$ we let the scrambler $\sigma: M \rightarrow M$ be a permutation on M . The definition of \models^σ is like that of \models except for:

$$m \models^\sigma \phi(t_1, \dots, t_n) \text{ iff } \langle \sigma(\mathfrak{I}(t_1)), \dots, \sigma(\mathfrak{I}(t_n)) \rangle \in \mathfrak{I}(\phi).^{13}$$

Truth-in-a-model becomes a special case of scrambled-truth-in-a-model when σ is the identity function.

Claim: For any sentence S of L , $m^L \models S$ iff $m' \models^{f^{-1}} S$.

Proof: The only cases to consider are the atomic sentences. Once the claim is established for those, extension to the general case by induction on complexity is routine. From our definitions,

$$\begin{aligned} m' \models^{f^{-1}} \phi(t_1, \dots, t_n) &\text{ iff} \\ &\langle f^{-1}(\mathfrak{I}'(t_1)), \dots, f^{-1}(\mathfrak{I}'(t_n)) \rangle \in \mathfrak{I}'(\phi) \text{ iff} \\ &\langle f^{-1}(f(\mathfrak{I}^L(t_1))), \dots, f^{-1}(f(\mathfrak{I}^L(t_n))) \rangle \in \mathfrak{I}^L(\phi) \text{ iff} \end{aligned}$$

$$\langle \mathfrak{S}^L(t_1), \dots, \mathfrak{S}^L(t_n) \rangle \in \mathfrak{S}^L(\phi) \text{ iff}$$

$$m^L \models \phi(t_1, \dots, t_n). \quad \square$$

Remark 1: m^L and m' , while agreeing in all the assignments to the predicate letters, disagree in their assignments to the terms. (Recall that f is nontrivial.) It is indeterminate which of \mathfrak{S}^L and \mathfrak{S}' is intended, and so indeterminate which of m^L and m' is intended. Assuming that singular reference is modeled by the restriction of interpretation functions to constants, singular reference goes indeterminate even under strong Lewisian assumptions.

Remark 2: A similar argument can be given for languages with no constants by focusing on the usual provision for interpreting variables. Let L' be just like L except without the constants. Letting $\mathfrak{S}^{L'}$ agree with \mathfrak{S}^L on every sign in L' , we let $m^{L'}$ be $\langle M^L, \mathfrak{S}^{L'} \rangle$. Given any assignment s of values from M^L to the free variables v , by analogous considerations to those above for any open formula $\phi(v_1, \dots, v_m)$ of L' , $m^{L'} \models \phi(v_1, \dots, v_m)$ iff $m^{L'} \models^{f^{-1}} \phi(v_1, \dots, v_m)$, where s' is the composition $f \circ s$. By induction on syntactic complexity it can then be shown that for any sentence S of L' , $m^{L'} \models S$ iff $m^{L'} \models^{f^{-1}} S$. The significance of this particular detail will emerge in the next section.

Remark 3: The present argument differs from familiar indeterminacy arguments originating from Quine and Putnam, with familiar responses by Lewis, Devitt, and others. The other arguments take for granted that truth *per se* is to be modeled by truth-in-a-model and then proceed by permuting assignments to the non-logical vocabulary in a truth-in-a-model preserving way.¹⁴ Not so in the present case. The idea that sentential truth is to be modeled by truth-in-a-model is not sacrosanct, a fact exploited to drive the present threat of referential indeterminacy.

4. Interpretationist Replies

How would the interpretationist proceed here? The interpretationist begins with truth-conditions for whole sentences. The interpretation of our predicate letters is already presumed to be fixed and beholden to maximal naturalness. The challenge is to say what selects whether to be true for sentences is as of \models or as of $\models^{f^{-1}}$. If something already selects whether reference is as of \mathfrak{S}^L or as of \mathfrak{S}' , then starting with truth-conditions for sentences might determine whether to be true is as of \models or as of $\models^{f^{-1}}$.¹⁵ Alternatively, starting with truth-conditions for sentences, whether reference is as of \mathfrak{S}^L or as of \mathfrak{S}' will depend, in turn, on whether to be true is as of \models or as of $\models^{f^{-1}}$. To borrow an image from Davidson, starting with truth-conditions and trying to solve for what generates them is like trying to solve a single equation with two variables, reference and truth.

Interpretationism prioritizes truth over reference. Reference is understood as the assignment of semantic values to expressions in generating truth-conditions for sentences. The Lewisian interpretationist understands this to be constrained by considerations of naturalness in the assignment of semantic values to predicates,

assignments that carve nature at the joints. Crucially, how subsentential expressions are to be interpreted is beholden to objective joints in nature. But more fundamentally, how subsentential expressions are to be interpreted is constitutive of the expressions having their significance to begin with. The argument of the previous section suggests that such an approach leaves singular reference, the kind of reference with which we engage in everyday life in speaking of apples at the fruit stall or of individual cats, radically under-determined.

At this point we might consider a Quinean reaction. Hoping to allay concerns about indeterminacy in singular reference, the Quinean insists on the eliminability of singular terms. Dispensing with singular terms dispenses with whatever afflicts them. But here two points should be borne in mind. First, if singular reference is important *pre-theoretically*, then the Quinean attitude will fail to engage with the problem at hand in a satisfactory way. And it seems very important pre-theoretically: we seem to care deeply about what *in particular* we think and talk about in our thought and talk. Indeed, we seem to care about this even when truth and falsity are not at issue.¹⁶ Second, in light of Remark 2 above, even if at some level our language were entirely bereft of singular terms it would be indeterminate what makes it the case that a claim of the form $\exists x\psi x$, e.g., comes out true, what the semantic mechanism is via which such truth is accomplished. As noted above, given an assignment to the free variables we can consider an assignment that is the composition of the original assignment and the inverse of the scrambler. It is then easily shown by appealing to the relevant semantic clause governing existential quantification that $m^{L'} \models \exists x\psi x$ iff $m^{L'} \models^{f^{-1}} \exists x\psi x$. But the mechanism via which $\exists x\psi x$ is made true in $m^{L'}$ is that something in the domain $M^{L'}$, call it o , falls in the Lewisian interpretation of ψ . The mechanism via which $\exists x\psi x$ is made f^{-1} -scrambled-true in $m^{L'}$, by contrast, is that something in M^L potentially other than o —the f of o —has an image under f^{-1} that falls in the Lewisian interpretation of ψ . A Quinean attitude might proclaim this a distinction without a difference, but those of us who wish to maintain a realist attitude about semantic facts would consider at most one of these alternatives to be correct.¹⁷

As against the Quinean attitude of indifference, let us be reminded what things would be like if truth were better captured by $\models^{f^{-1}}$ than by \models . I say ‘This is a nice piece of fruit’ standing at the fruit stall in my neighborhood grocer. It so happens that ‘this’ as spoken by me as I consider a juicy Ambrosia apple in the stall really refers to the exoplanet Alpha Centauri Bb, some four light years away. It also happens that for my sentence to be true is for an image of the exoplanet under some bijection of the universe onto itself to be a nice piece of fruit. Lo and behold, the image in question is just the apple I am holding in my hand, which is a nice piece of fruit indeed—a fortuitous cosmic coincidence.

But again, ‘this’ is a singular term and our insistent Quinean proposes that we do without those. We adjust the example to accommodate the insistence. Looking through the stall I say ‘There is a nice piece of fruit here—I can just feel it’. Ignoring the treatment of ‘here’, my first sentence might be made true by the fact that something somewhere in the universe—exoplanet Alpha Centauri Bb—is such that its image under a certain bijection of the universe onto itself is a certain apple in

the stall here in my neighborhood grocer, four light years away from the exoplanet, which happens to be a nice piece of fruit. This should strike us as no less absurd than the outlandish alternative outlined with respect to ‘This is a nice piece of fruit’. If I am right, the interpretationist has no obvious way of doing justice to our inclination to regard such alternatives as wildly implausible.

Might the interpretationist appeal to some general consideration of simplicity favoring truth-in-a-model over scrambled-truth-in-a-model, thereby allowing us to weed out as unintended certain overall interpretations of our language that are maximally eligible with respect to the predicates? After all, the definition of scrambled-truth-in-a-model, with its appeal to a scrambler σ , seems more complicated than that of truth-in-a-model. Unfortunately, such a response is of limited reach. For one thing, truth-in-a-model can be construed as a special case of scrambled-truth-in-a-model, as noted above, where the scrambler is identity. And we would be hard pressed to find a non-*ad-hoc* way of regarding identity as somehow ‘simpler’ than f^{-1} . The challenge is to spell out a way in which the identity function, all on its own, is simpler than an arbitrary nontrivial permutation of the domain. Against the background of comparing permutations of the domain, identity is not obviously simpler: it is one permutation among many, but one that requires a further condition to specify. But then again, against such a background identity is not being considered ‘all on its own’.¹⁸ How to think of the comparative simplicity of identity but not against any such background is unclear.¹⁹

One might appeal to some formally desirable property that truth-in-a-model possesses and scrambled-truth-in-a-model lacks, such as invariance under isomorphism.²⁰ The following expresses the requirement for truth-in-a-model:

- (i) If $m \models S$, then for any $m^* = \langle M^*, \mathfrak{S}^* \rangle$ isomorphic to m , $m^* \models S$.

The strict analog of (i) for scrambled-truth-in-a-model fails because the scrambler σ is a permutation on M and there is surely no guarantee that for every term t and every such m^* , $\mathfrak{S}^*(t) \in M$. So there is no guarantee that $\sigma(\mathfrak{S}^*(t))$ is well defined. But the relevance of the requirement of invariance under isomorphism to the supposed advantage of truth-in-a-model over scrambled-truth-in-a-model for modeling truth *per se* is not obvious. Let us assume that under the auspices of abstract model theory (i) has a clear advantage over the scrambled variant (ii):

- (ii) If $m \models^\sigma S$, then for any $m^* = \langle M^*, \mathfrak{S}^* \rangle$ isomorphic to m , $m^* \models^{\sigma^*} S$, where $\sigma^* \text{ is } h \circ \sigma \circ h^{-1}$ and $h: M \rightarrow M^*$ is the isomorphism.²¹

Such an advantage of (i) over (ii) from the point of view of abstract model theory does not immediately entail without further argument a clear advantage for truth-in-a-model over scrambled-truth-in-a-model as the better capture of truth *per se* for our sentences. And such an argument seems like a very tall order indeed.

But even if we set all this aside, compare the suggested appeal to simplicity with the following possible response—not attempted by Putnam—to the Lewisian appeal to reference magnetism as the saving constraint on the intended interpretation of the language of the epistemically ideal theory T . The idea that to be true for T is to be true in some model with the world itself as the domain is certainly *locally* simpler

than the idea that to be true for T is to be true in a model with the world itself as the domain and with an interpretation function that respects the world's preexisting structure.²² Thus, a blind appeal to simplicity would drive the shocking conclusion that T is guaranteed to come out true of the world after all, so that the distinction between epistemic ideality and realist truth cannot be sustained. But such an appeal to simplicity is clearly otiose. The question posed by Putnam's argument is whether or not T could really be false. Such an issue has ramifications that are far-reaching enough to trump any knee-jerk appeal to simplicity that would favor the locally simpler account of what it is for T to be true.

In the present case, too, if singular reference is as of \mathfrak{S}' rather than as of \mathfrak{S}^L , then this provides us with ample reason to set aside a consideration of simplicity that would favor truth-in-a-model (or identity-scrambled-truth-in-a-model) over f^{-1} -scrambled-truth-in-a-model as the better capture of truth for our sentences. After all, when I say 'This is a nice piece of fruit' while attending to the apple in my hand, it had better turn out that what I say is true or not depending on how things stand in the vicinity of the apple. If 'this' as spoken at the fruit stall really refers to Alpha Centauri Bb, then blindly following some local simplicity criterion would make what I say true or false depending on how things stand with something other than the apple some four light years away.

The interpretationist might try to shift somehow to a consideration of simplicity regarding reference and truth taken together. But notice that the matter is not as clear as one might have hoped. We seem to have no grounds for supposing that singular reference, taken on its own, is simpler one way or another: the hypothesis that reference is as of \mathfrak{S}' is no less simple than the hypothesis that reference is as of \mathfrak{S}^L . And as seen above, we seem to have no reason to suppose that simplicity considerations automatically trump when it comes to truth-in-a-model (or identity-scrambled-truth-in-a-model) over f^{-1} -scrambled-truth-in-a-model. How exactly simplicity considerations are supposed to trump when it comes to reference and truth taken together is far from clear.

Perhaps the thought is that Lewisian considerations of naturalness may be taken to rule out scrambled-truth-in-a-model in favor of truth-in-a-model directly.²³ But naturalness as conceived by Lewis does not apply smoothly here: we lack a workable sense of what the equivalent of natural joints might be outside the natural order. Lewis (1983: 375–6) does employ the apparatus of natural properties to solve the Kripke-Wittgenstein problem, but here care must be taken not to draw unintended conclusions. The question posed by Kripke's Wittgenstein is what determines that by adding we mean to add rather than to quadd. Lewis's answer is that the property of adding is more natural than the property of quadding. This should not be mistaken for the claim that the function of addition is more natural than the function of quaddition and that therefore 'addition' refers to the one and not the other. In the first place, a function is not a property but an individual, and while Lewis does have something to say about eligibility for individuals in terms of demarcation of the individual's boundaries by more natural properties (as discussed in Section 2), functions clearly lack boundaries in the relevant sense. Such an account for mathematical individuals would need a notion of naturalness for mathematical

properties that Lewis does not provide. For some indication of the difficulty here, consider which is more natural in the relevant sense—being a natural number or being a real? Since every natural is a real and not the other way, the naturals are certainly more exclusive, which might suggest that they are more elite in the relevant sense. But by that criterion the transcendentals are also more exclusive than the reals—are they, too, more elite? Or is it that being of the same cardinality as the reals precludes the transcendentals from being more elite? The point is not that we could not dream up a criterion come what may. The point is that such work is far from trivial and clearly not a straightforward extension of the Lewisian apparatus of naturalness. The relevant context in Lewis (1983) makes clear that Lewis's focus is on naturalness in the psychology of adding as compared with the psychology of quadding. The focus is not on naturalness in interpreting mathematical language, appealing somehow to the comparative naturalness of the addition function. The focus is on naturalness in interpreting the attitudes undergirding the mental activity of adding. The various cases of adding are supposed to enjoy greater objective similarity than the various cases of quadding. In sum, attempting to rule out scrambled-truth-in-a-model by appealing to Lewisian naturalness is problematic. There is no straightforward extension of the Lewisian apparatus to the case at hand.

Let me conclude this section by considering one last interpretationist response to the above argument that relies on the role reference plays in the overall explanation of why sentences have their truth-conditions.²⁴ The response builds on an idea found in Sider (2011):

Following J. Robert G. Williams (2007, section 2) we can derive the doctrine of reference magnetism from a well-motivated and more general doctrine about theoretical virtue. This doctrine is the one defended in section 3.1: explanatory theories must be cast in joint-carving terms.

As I will develop it, the crucial assumption of the derivation is that *reference is an explanatory relation*—one can explain certain facts by citing what words refer to. But if reference were given a bizarre interpretation, then reference-involving 'explanations' would not in fact be explanatory, since they would be cast in badly non-joint-carving terms. (27–8)

The interpretationist wants to say that part of the reason why 'This is a piece of fruit' has its truth-conditions is that 'this' refers to the apple. The fact that 'this' refers to the apple plays a certain role in the explanation of semantic endowment for the sentence. But this seems not so under the deviant interpretation that has 'this' refer to the planet. For under the proposed alternative scheme it does not matter *what* 'this' refers to as long as the scrambler maps it onto the apple. In this way truth-conditions for 'This is a piece of fruit' seem insensitive to whatever 'this' refers to. It would seem to preclude reference from playing its requisite explanatory role.

The reply to the interpretationist is that under the f^{-1} -scrambled-truth-in-a-model scheme the truth-conditions for 'This is a piece of fruit' are just as sensitive to whatever 'this' happens to refer to. Nothing precludes the reference relation from playing a critical explanatory role in why sentences have their truth-conditions. Indeed, if sentential truth is as of f^{-1} -scrambled-truth-in-a-model, then the reference

relation that takes ‘this’ to the planet rather than the apple has as much of an explanatory role to play in the association of sentences with their truth-conditions as the reference relation that takes ‘this’ to the apple has if sentential truth is as of identity-scrambled-truth-in-a-model. Whether sentential truth is one way while reference is as of \mathfrak{S}^L , or sentential truth is another way while reference is as of \mathfrak{S}' , the allocation of truth-conditions to sentences will be the same (*modulo* the ambiguity about truth). Sentential truth being the second way does not preclude singular reference (as modeled by \mathfrak{S}') from playing its explanatory role if we assume that sentential truth being the first way does not preclude singular reference (as modeled by \mathfrak{S}^L) from playing its explanatory role. In both cases the reference of ‘this’ partially explains why the sentence has its truth-conditions. The explanatory role of reference under the f^{-1} -scrambled-truth-in-a-model scheme is in no way diminished.

5. Productivism Redux

Things are different under the auspices of metasemantic productivism. Productivism prioritizes reference over truth.²⁵ What determines semantic significance for subsentential expressions are conditions surrounding their production and manipulation by speakers or writers. Crucially, their significance is fixed prior to raising and settling questions about how they should be interpreted. Productivism can thus directly rule out m' or m^L (or both) as unintended—whichever gets antecedently produced reference wrong. Reference is determined by the conditions surrounding the production of the referring items, which allows us to select interpretations as intended and discard others as unintended.²⁶ For example, on a referential-intention-based productivist story, ‘this’ as spoken with respect to a particular apple depends for its semantic endowment on the speaker’s intention to refer to the apple, which plausibly depends, in turn, on a causal-historical link of the right sort obtaining between speaker and apple. By contrast, no such referential intentions exist to back up ‘this’ as produced on the relevant occasion referring to Alpha Centauri Bb or to the number 17.²⁷

Productivism also has resources, which interpretationism lacks, to explain directly why sentential truth is better modeled by truth-in-a-model (the limit case of scrambled-truth-in-a-model where the scrambler is identity) than by scrambled-truth-in-a-model with no restriction on the scrambler. Sentential truth is local as per produced reference in a way that scrambled-truth-in-a-model with no restriction on the scrambler ignores. Locality in this context should be thought of as direct dependence. If ‘This is a piece of fruit’ expresses a truth, then given that the reference of ‘this’ is already to the apple in the stall, we naturally and intuitively require that the truth in question should turn directly on how things stand with the apple in question, rather than turn directly on how things stand with an individual other than the apple instead. In the limit case of scrambled-truth-in-a-model where the scrambler is identity this intuitive requirement of locality-per-reference on truth is clearly respected.²⁸ This is why we consider truth-in-a-model a better theoretical capture of sentential truth than scrambled-truth-in-a-model with no restriction on the scrambler. Under the auspices of productivism, reference is settled and *decides*

the interpretation of singular terms. We are then in a good position to appeal to an intuitive requirement on truth—locality-per-reference—to argue against the suitability of scrambled-truth-in-a-model. Productivism is at a clear advantage here. As we saw, for an interpretationist locality considerations might deliver the wrong result, given that reference is not antecedently settled. If singular reference is as of \mathfrak{S}' rather than as of \mathfrak{S}^L , then in the interest of maintaining the distribution of truth-conditions over sentences the better capture of truth could easily violate the locality-per-reference requirement. Not so under the auspices of productivism.²⁹

Let me conclude by highlighting the sheer intuitive plausibility of a productivist outlook in metasemantics as compared with interpretationism. Metasemantic interpretationism is a surprising doctrine easily mistaken for a benign and uncontroversial one. The doctrine maintains that the significance of expressions is *constituted* by their interpretability, whether by an actual linguistic actor or by an idealized version thereof. Talk of constitution emphasizes the distinctly metaphysical flavor of the view, that it is a metasemantic doctrine targeting the creation of endowment with significance, the conditions surrounding its emergence. Such talk serves to distinguish the view as discussed here from a mild and rather plausible doctrine in the epistemology of understanding. There can be no serious question as to whether interpretation plays a crucial role when it comes to semantic uptake. How else might we come to appreciate the significance of demonstrative pronouns, say, if not by constructing interpretations, understood as empirical conjectures of a sort? The conjectural aspect of such an endeavor is brought into sharp relief in cataphoric contexts, cases where the audience has to keep track and backtrack to earlier portions of the discourse in light of what happens later. This all concerns the epistemology of understanding. Metasemantic interpretationism, on the other hand, is not an epistemological doctrine but a metaphysical one. The question it sets out to answer is not how we come to know what expressions mean, but, rather, how it is that they mean what they do, what it is that confers significance upon them. A speaker begins ‘She was unsure at first’ and then stops short, for whatever reason, of completing the utterance with ‘but then Amy made up her mind to leave’.³⁰ Let us assume that the utterance is sudden enough and out of the blue to leave the audience in the dark as to the significance of ‘she’. Two metasemantic questions immediately arise. First, does ‘she’ as spoken on that occasion succeed in referring to anyone in particular? Second, assuming that ‘she’ does succeed in this way, what makes it the case that it stands for Amy rather than for anyone else? Productivism typically gives an affirmative answer to the first question. And it would typically appeal, as part of its answer to the second question, to facts concerning the history of the speaker vis-à-vis Amy that enter into the production of the token on the relevant occasion: for example, that it was Amy that the speaker had in mind as the intended referent for the produced token of ‘she’. Interpretationism might give an affirmative answer to the first question as well. But as an answer to the second question the interpretationist would appeal to features of the larger discourse in which the token embeds and the availability of a global interpretation of that discourse that maximizes speaker rationality and truth given the circumstances of the speaking.

My aim in this paper has been to demonstrate that interpretationism is confronted with a distinctive challenge concerning singular reference that the Lewisian doctrine of reference magnetism is not equipped to handle. The problem raised pivots on the nature of being true for sentences. But extending reference magnetism to the predicate ‘true’ is of little use. It can be agreed on all sides that ‘true’ as applied to sentences stands for being true as applied to sentences. The threat of indeterminacy discussed in this paper proceeds by targeting the nature of being true for sentences, not by targeting the connection between the predicate ‘true’ and being true. And reference magnetism is silent on the nature of sentential truth.

There remains the possibility of future extensions and elaborations of the Lewisian outlook in light of the above, supplementing reference magnetism with a direct engagement with the notion of sentential truth.³¹ But productivism already goes a certain distance towards engagement with the issue, and without the unnatural subjugation of reference to truth. Metasemantic productivism is easily and naturally informed by an intuitive requirement that truth for sentences be local-per-reference. Language-world relations pertaining to syntactically complex expressions of a certain sort turn out to depend on language-world relations pertaining to syntactically simpler constituents rather than the other way around. But for interpretationist reconstructions to the contrary, this is exactly how things should be.

I began this paper by noting the obvious fact that metasemantics is beholden to semantics. Semantics studies the *what* of semantic endowment while metasemantics studies the *how*. But having come this far we can now appreciate that at a different level semantics can be seen as beholden to metasemantics as well. Through the comparison with scrambled-truth-in-a-model, truth-in-a-model is seen as an obvious natural choice for a basic semantic notion.³² The historical metamathematical development of truth-in-a-model in the work of Tarski, Vaught, and others, should not blind us to the fact that utilizing the notion for natural language semantics is particularly compelling—in a way that utilizing scrambled-truth-in-a-model with no restriction on the scrambler could never be. This is so to the extent that model-theoretic semantics is already tacitly committed to the natural and intuitive requirement that truth be local-per-reference. This locality requirement, I have argued, sits ill with a metasemantics whereby whole sentences and their truth-conditions are earlier in the order of metasemantic explanation than the reference of singular terms, the latter being a subsequent abstraction. If I am right, model-theoretic natural language semantics, with its formal articulation of truth’s locality-per-reference, already exhibits a tacit commitment to a non-interpretationist metasemantics.

Notes

*For helpful discussion of this material I am indebted to audiences at UBC, University of Bologna, Bar Ilan University, UC Santa Cruz, Lund University, Tel Aviv University, Princeton University, University of Alberta, and to the following individuals: Mahrad Almotahari, Roberta Ballarin, Paul Bartha, Yuval Dolev, Eli Dresner, Delia Graff Fara, Liz Harman, Robert Hirsch, Emmanuel Genot, Justine Jacot, Tom Kelly, Kathrin Koslicki, James Martin, Chris Mole, Howard Nye, Erik Olsson, Alan Richardson, Gideon Rosen, Chris Stephens, Abe Stone, Max Weiss, and two anonymous referees. I gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada.

¹Semantic endowment may be multi-layered. The type ‘this’ is governed by a certain semantic rule—a character—that specifies semantic values—contents—of tokens of the type in their respective contexts. One metasemantic issue that could be raised in light of this semantic proposal is how the type ‘this’ came to be associated with its character. A different metasemantic issue is how a given token of ‘this’ came to be associated with its particular contribution to truth-conditions, the individual demonstratum. In what follows I focus exclusively on metasemantic issues of the second kind.

²Cf. Wittgenstein (1961), at 3.32: “The sign (‘Zeichen’) is what can be perceived of a symbol.” Extending ‘perceive’ to cover introspection, the distinction can be extended to cover mental items as well.

³This does not preclude—and most often includes—the requirement that the agent be appropriately causally linked to the individual the referential intention specifies. For further discussion of such details regarding names, see Chapter 3 of Simchen (2012).

⁴Whether or not causal-historical rapport with a morphological item is sufficient for causal-historical rapport with an intended referent need not concern us. I also note in passing that a teleosemantic account whereby semantic endowment arises from the diachronically shaped function of the produced item to represent to the item’s consumers is a different sort of productivist account. For discussion of the role of consumers here, see Millikan (1989) and especially (1990).

⁵For interpretationism the item’s regardability under various constraints determines its semantic endowment directly. This should be distinguished from a productivist alternative whereby the item’s semantic endowment is shaped by intentions on the side of the producer that include the intention that the item be regarded in a certain way. On the latter productivist alternative the item’s regardability enters into determining semantic endowment only indirectly, via conditions of production that include the relevant intention on the side of the producer. Here is a recent endorsement of such an idea (not for the purpose of promoting any particular metasemantic position) in Heck (2014):

Successful communication requires the speaker and her audience to converge on a referent. But the speaker does not utter the demonstrative and then consult the contextual cues to figure out how to interpret her own words. Rather, in planning her speech, she has already decided what object to assign as value of the contextual parameter that fixes the referent of the demonstrative, that is, which object she intends her audience to interpret her as speaking about. (343)

And here is a recent endorsement of the idea that potential uptake by a cooperative audience is to be taken into account within a broadly intention-based productivist account of demonstrative reference in King (2014):

I suggest we say that the value of a use of a demonstrative in a context is that object *o* that meets the following two conditions: 1) the speaker intends *o* to be the value; and 2) a competent, attentive, reasonable hearer would take *o* to be the object that the speaker intends to be the value. We can abbreviate this by saying that an object *o* is the value of an occurrence of a demonstrative in context just in case the speaker intends *o* to be the value and the speaker successfully reveals her intention. [Here a footnote is added: “Note that a speaker can successfully reveal her intention even though her hearer failed to figure out what she intended. The hearer could be inattentive, incompetent, etc.”—OS] (225)

⁶The approach has been extended by Sider (2011) to cover assignments of semantic values to expressions other than predicates, notably quantifiers, but this extension does not affect the main focus here, which is singular reference, beyond Lewis’s efforts under the original proposal.

⁷For present purposes we set aside important questions regarding Lewis’s metasemantic orientation that matter a great deal to Lewis scholarship. See, in particular, Weatherston (2012) and Schwartz (2014). In Lewis (1984) we come across the following caveat:

I shall acquiesce in Putnam’s linguistic turn: I shall discuss the semantic interpretation of language rather than the assignment of content to attitudes, thus ignoring the possibility that the latter settles the former. It would be better, I think, to start with the attitudes and go on to language. But I think that would relocate, rather than avoid, the problem; wherefore I may as well discuss it on Putnam’s own terms. (222)

In everything that follows we acquiesce in Lewis’s acquiescence.

⁸We may steer away from controversy surrounding the implied suggestion that the way the world really is is a model by paraphrasing as follows: T would be false if it happens not to have a model isomorphic to a model representing the way the world really is.

⁹Lewis credits Merrill (1980) with the general idea.

¹⁰Thus we read in Putnam (1990): “What Lewis’s story claims is that the class of cats cries out for a label, while the class of cats* does not cry out to be named. Rather than solving the problem of reference, what the idea of a constraint built into nature and of ‘elite classes’ does is to confuse the materialist picture by throwing in something ‘spooky.’” (38)

¹¹As should be clear from the ensuing discussion, the example of demonstrative reference is selected as a way of dramatizing indeterminacy in singular reference. Nothing important hangs on distinctive features of demonstrative reference beyond utility for singular reference. Any referentially used singular term would do just as well for present purposes, such as a referential use of ‘Bruce’ to speak of a particular cat.

¹²Our discussion concerns Lewis’s antidote to Putnam-style referential indeterminacy in an extensional setup, so what is held fixed is the assignment to the predicate letters of so-called elite classes—extensions of natural properties and relations—rather than the properties and relations themselves. Referential indeterminacy arguments can be tailored to suit richer languages as well. See Chapter 2 of Putnam (1981) and Hale and Wright (1997).

¹³We assume that where the terms are variables the definition relativizes to an assignment s of values to the variables:

$$m, s \models^{\sigma} \phi(v_1, \dots, v_n) \text{ iff } \langle \sigma(s(v_1)), \dots, \sigma(s(v_n)) \rangle \in \mathfrak{N}(\phi).$$

¹⁴For details see Hale and Wright (1997). Taking for granted that truth *per se* is to be modeled by truth-in-a-model applies equally to arguments that proceed by construction of deviant interpretations that are arguably simpler than the intended one and to arguments that conceive of alternative semantic properties and relations that differ from one another in their patterns of instantiation. See Williams (2007) and Hawthorne (2007).

¹⁵‘Might’ because differences among bijections may not be significant enough to be linguistically manifestable, in which case the resolution alluded to in the text would not work. Henceforth we restrict ourselves to differences that are *prima facie* linguistically manifestable.

¹⁶See Simchen (2013) for discussion of cases of truncated utterances that are referential despite having only subsentential significance. (One may think of such cases as playing a non-negligible communicative role in bringing topics into conversational salience.)

¹⁷A discussion of variable elimination, as in Quine (1960) and elsewhere, would take us too far afield, but predicate-functor logic, with its *derelativization* (or *cropping*) functor, seems too syntactically alien to play a significant modeling role in natural language semantics. For a clear discussion of the issue that also includes a straightforward model theory for PFL, see Dahllöf (1999).

¹⁸It might be tempting to appeal to degrees of Kolmogorov complexity here, but notice that the relevant issue is not which description of a function is simpler, but, rather, which function is simpler. It is the latter rather than the former that is relevant to the question at hand of whether or not truth-in-a-model (or identity-scrambled-truth-in-a-model) is simpler than f^{-1} -scrambled-truth-in-a-model. How we are to proceed from computational simplicity in specifying things to simplicity in the things themselves remains unclear. (For what it’s worth, mathematicians’ verdicts of simplicity are often shaped by how much is required to specify things, so e.g. a monoid might be considered simpler than a group due to the absence of the inverse condition. By such a standard a non-arbitrary permutation will not be simpler than an arbitrary one.)

¹⁹In the spirit of the previous discussion of the Quinean attitude of indifference it might be claimed that truth-in-a-model and identity-scrambled-truth-in-a-model are after all distinct, the latter but not the former operating via the workings of the identity function. The point deserves a more thorough treatment than I can offer it here, but suffice it to say that once the comparison with the neighboring scrambled notion is brought into salience, it becomes unclear whether the former construal should be privileged over the latter. Thanks to James Martin for raising this issue.

²⁰Thanks to Max Weiss for emphasizing this.

²¹*Claim*: Scrambled-truth-in-a-model has (ii).

Proof: We show that if $m \models^{\sigma} \phi(t_1, \dots, t_n)$, then for any $m^* = \langle M^*, \mathfrak{S}^* \rangle$ isomorphic to m , $m^* \models^{\sigma^*} \phi(t_1, \dots, t_n)$, where σ^* is $h \circ \sigma \circ h^{-1}$ and $h: M \rightarrow M^*$ is the isomorphism. Extension to the general case by induction on syntactic complexity is again routine. For $m = \langle M, \mathfrak{S} \rangle$, $m \models^{\sigma} \phi(t_1, \dots, t_n)$ iff $\langle \sigma(\mathfrak{S}(t_1)), \dots, \sigma(\mathfrak{S}(t_n)) \rangle \in \mathfrak{S}(\phi)$ iff

$$(\dagger) \langle h(\langle \sigma(\mathfrak{S}(t_1)), \dots, \sigma(\mathfrak{S}(t_n)) \rangle) \rangle \in \mathfrak{S}^*(\phi).$$

On the other hand, for any term t , $\mathfrak{S}^*(t) = h(\mathfrak{S}(t))$, so $\mathfrak{S}(t) = h^{-1}(\mathfrak{S}^*(t))$. Substituting in (\dagger) yields:

$$(\dagger') \langle h(\sigma(h^{-1}(\mathfrak{S}^*(t_1))), \dots, h(\sigma(h^{-1}(\mathfrak{S}^*(t_n)))) \rangle \in \mathfrak{S}^*(\phi).$$

We observe that $h \circ \sigma \circ h^{-1}$, i.e. σ^* , is a permutation on M^* , so (\dagger') obtains iff $m^* \models^{\sigma^*} \phi(t_1, \dots, t_n)$ by the definition of scrambled-truth-in-a-model. \square

²² ‘Locally’ because the idea of the world being a totality of things with no inherent structure, what with minds imposing structure on this totality, may very well turn out to have ramifications of enormous complexity on a global scale.

²³ Thanks to an anonymous referee for raising this.

²⁴ Thanks to Mahrad Almotahari for discussion here.

²⁵ It might be thought that the contrast between productivism and interpretationism does not coincide with a contrast between allotting explanatory priority to semantic endowment for subsentential expressions as opposed to allotting it to semantic endowment for full sentences (as is the wont of both Davidsonian and Lewisian varieties of interpretationism). This might be so as a purely conceptual matter, but considerations I offer in Simchen (2013) tell against the option of a productivist account that allots explanatory priority to semantic endowment for full sentences. And familiar Quinean considerations regarding the evidential situation for interpretation deem an interpretationist account that allots explanatory priority to semantic endowment for subsentential expressions unpromising as well. See, e.g., Quine (1968) and Davidson (1977). Thanks to Eli Dresner and Liz Harman for discussion.

²⁶ Even though our focus here is on singular terms, this is no less true for predicates. See Simchen (2015) for important ramifications of this point for interpretation in general and in legal contexts in particular.

²⁷ How to think of referential intentions and other *de re* attitudes directed at numbers is vexing. Suggestive work on the topic within a broadly productivist outlook can be found in Kripke (1992).

²⁸ Here is a case that violates the intuitive requirement of locality-per-reference on truth (where such locality is, once again, a matter of direct dependence): Suppose that truth is determined by microphysical goings on and yet singular reference is invariably to macro things. On such a view, what t refers to—a macro thing—is not identical (but merely coincides, let us say) with that whose falling under ϕ makes ϕt come out true—a molecular lattice. While there is a sense of ‘local’ according to which truth in such a case still abides by the intuitive requirement—the macro thing and the lattice are at the same place at the same time—that is not the sense relevant for the requirement.

²⁹ The committed interpretationist might try to argue that the advantage just discerned for productivism over interpretationism is illusory. Productivism, it might be claimed, is just another theoretical capture, a metasemantic *theory*, to which reference magnetism applies in turn. Some Lewisians are certainly drawn to such a ‘just more metasemantic theory’ move against a productivist orientation, shifting from reference-magnetism in metasemantics to what Sider (2011) calls *metametaseantics*. The matter deserves a separate discussion that cannot be undertaken here, but see Chapter 1 of Simchen (unpublished).

³⁰ For further discussion of such phenomena and their relevance for metasemantics, see Simchen (2013).

³¹ It might be argued on essentialist grounds that it is of the very nature of sentential truth that it is better modeled by truth-in-a-model than by scrambled-truth-in-a-model. But if the foregoing is correct, metasemantic productivism allows us to avoid such an independent heavy-duty essentialist commitment regarding the nature of sentential truth. Under the auspices of productivism the requirement that truth be local-per-reference is natural and highly intuitive by comparison. Thanks to Gideon Rosen for discussion here.

³² Lest it be suspected that the issue raised in this paper is peculiar to model-theoretic semantics while a truth-theoretic approach is somehow immune to it, I offer an adaptation of the situation to a truth-theoretic setting in the Appendix below.

³³ See Davidson (1974).

³⁴ Corner quotes are utilized throughout to minimize use-mention confusion.

Appendix: Scrambled Truth

We show the availability of a notion of scrambled truth in a truth-theoretic setting by considering a first-order extensional toy language L^H . Interpretation begins with empirical hypotheses about L^H in the form of T-sentences.³³ Say we have amassed the following T-sentences, among others:

- $\ulcorner \text{SOCRATES ADAM O KELEV} \urcorner$ is true iff Socrates is human or Socrates is a dog.³⁴
- $\ulcorner \text{FIDO LO ADAM} \urcorner$ is true iff Fido is not human.
- $\ulcorner \text{MASHEHU ADAM VE LO KELEV} \urcorner$ is true iff something is human and not a dog.
- $\ulcorner \text{SOCRATES KELEV O MASHEHU ADAM VE KELEV} \urcorner$ is true iff either Socrates is a dog or something is both human and a dog.

Our aim is to give a definition of the truth predicate that entails these. The first interpretive task is to discern semantically significant units within them. Say we conclude that $\ulcorner \text{SOCRATES} \urcorner$ and $\ulcorner \text{FIDO} \urcorner$ are terms, $\ulcorner \text{ADAM} \urcorner$ and $\ulcorner \text{KELEV} \urcorner$ are one-place predicates, and the logical particles are as follows: the $\ulcorner \text{MASHEHU} \urcorner$ construction is existential quantification, the $\ulcorner \text{VE} \urcorner$ construction is conjunction, the $\ulcorner \text{O} \urcorner$ construction is disjunction, and the $\ulcorner \text{LO} \urcorner$ construction is negation. So the logical form of $\ulcorner \text{SOCRATES ADAM O KELEV} \urcorner$ is rendered more explicit by unpacking the disjunction: $\ulcorner \text{SOCRATES ADAM O SOCRATES KELEV} \urcorner$. And the logical form of $\ulcorner \text{MASHEHU ADAM VE LO KELEV} \urcorner$ is rendered more explicit by adding a variable to be bound by the quantifier: $\ulcorner \text{MASHEHU } \nu \nu \text{ ADAM VE LO } \nu \text{ KELEV} \urcorner$. We thus add variables to the inventory of terms, $\ulcorner x_i \urcorner$ for each i .

Next we have the semantic clauses for denotation of a term relative to a sequence and application of a predicate. For any sequence s and any i , $\ulcorner x_i \urcorner$ denotes o relative to s iff o is the i^{th} member of s . For any sequence s and name n , n denotes o relative to s iff either $n = \ulcorner \text{SOCRATES} \urcorner$ and $o = \text{Socrates}$, or else $n = \ulcorner \text{FIDO} \urcorner$ and $o = \text{Fido}$. Finally, for any predicate P , P applies to o iff either $P = \ulcorner \text{ADAM} \urcorner$ and o is human, or else $P = \ulcorner \text{KELEV} \urcorner$ and o is a dog.

Next comes an inductive definition of satisfaction by a sequence. For any formula F of L^H and any sequence s drawn from the domain U that L^H is used to talk about, F is satisfied by s iff either 1. F is an atomic formula $\ulcorner tP \urcorner$ where t denotes o relative to s and P applies to o ; or else 2. F is a formula $\ulcorner \text{LO } G \urcorner$ for some formula G and G is not satisfied by s ; or else 3. F is a formula $\ulcorner G \text{ VE } H \urcorner$ for some formulas G and H and both G and H are satisfied by s ; or else 4. F is a formula $\ulcorner G \text{ O } H \urcorner$ and either G is satisfied by s or H is satisfied by s ; or else 5. F is a formula $\ulcorner \text{MASHEHU } x_i \text{ } G \urcorner$ where G has $\ulcorner x_i \urcorner$ free and there is a sequence s^* that differs from s in at most the i^{th} place such that G is satisfied by s^* . Finally: a sentence S of L^H is true iff for any sequence s of U , S is satisfied by s .

We now define scrambled truth as follows. Let μ be some permutation on U such that $\mu(\text{Socrates}) = \text{Fido}$ and $\mu(\text{Fido}) = \text{Socrates}$. For any sequence s and any i , $\ulcorner x_i \urcorner$

scrambledly denotes $\mu(o)$ relative to s iff o is the i^{th} member of s . For any sequence s and name n , n *scrambledly* denotes o relative to s iff either $n = \ulcorner \text{SOCRATES} \urcorner$ and $o = \text{Fido}$, or else $n = \ulcorner \text{FIDO} \urcorner$ and $o = \text{Socrates}$. For any predicate P , P applies to o iff either $P = \ulcorner \text{ADAM} \urcorner$ and o is human, or else $P = \ulcorner \text{KELEV} \urcorner$ and o is a dog, as before.

Next comes the inductive definition of scrambled satisfaction by a sequence, the only difference from satisfaction by a sequence being the first clause: for any term t and predicate P , $\ulcorner tP \urcorner$ is *scrambledly satisfied* by s iff t scrambledly denotes o relative to s and P applies to $\mu^{-1}(o)$. A sentence S of L^H is *scrambledly true* iff for any sequence s of U , S is scrambledly satisfied by s .

Claim: For any sentence S of L^H , S is true iff S is scrambledly true.

The proof is obvious from the definitions but tedious. Let us illustrate, however, with respect to a couple of sentences, first an atomic sentence and then one that is syntactically more complex.

First we show that for any sequence s of U , $\ulcorner \text{SOCRATES KELEV} \urcorner$ is satisfied by s iff $\ulcorner \text{SOCRATES KELEV} \urcorner$ is scrambledly satisfied by s . $\ulcorner \text{SOCRATES KELEV} \urcorner$ is scrambledly satisfied by s iff $\ulcorner \text{KELEV} \urcorner$ applies to the image under μ^{-1} of the scrambled denotation of $\ulcorner \text{SOCRATES} \urcorner$ relative to s , i.e. the image under μ^{-1} of Fido, i.e. Socrates. So $\ulcorner \text{SOCRATES KELEV} \urcorner$ is scrambledly satisfied by s iff Socrates is a dog, which holds iff $\ulcorner \text{SOCRATES KELEV} \urcorner$ is satisfied by s .

Next we show that for any s of U , $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is satisfied by s iff $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s . First, s satisfies $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ iff there is a sequence s^* that differs from s in at most the 17th place such that $\ulcorner x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is satisfied by s^* . The latter holds iff both $\ulcorner x_{17} \text{ ADAM} \urcorner$ and $\ulcorner \text{LO } x_{17} \text{ KELEV} \urcorner$ are satisfied by s^* , which holds iff $\ulcorner x_{17} \text{ ADAM} \urcorner$ is satisfied by s^* and $\ulcorner x_{17} \text{ KELEV} \urcorner$ is not satisfied by s^* , which holds iff something in U is human and not a dog. We now show that something in U is human and not a dog iff $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s . Suppose, first, that something in U is human and not a dog and assume for *reductio* that $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is not scrambledly satisfied by s . Then for any sequence s' that differs from s in at most the 17th place, $\ulcorner x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is not scrambledly satisfied by s' . So for any such sequence s' , either $\ulcorner x_{17} \text{ ADAM} \urcorner$ is not scrambledly satisfied by s' , or else $\ulcorner \text{LO } x_{17} \text{ KELEV} \urcorner$ is not scrambledly satisfied by s' so that $\ulcorner x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s' . So for any such sequence s' , either $\ulcorner \text{ADAM} \urcorner$ does not apply to the image under μ^{-1} of the scrambled denotation of $\ulcorner x_{17} \urcorner$ relative to s' , which is just the occupant of the 17th place in s' , or else $\ulcorner \text{KELEV} \urcorner$ applies to that occupant. This implies that everything in U is either not human or a dog, contradicting our assumption that something in U is human and not a dog. Therefore, if something in U is human and not a dog, then $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s . Finally, if $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s , then for some

sequence s^{**} that differs from s in at most the 17th place, $\ulcorner_{x_{17}} \text{ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s^{**} , so both $\ulcorner_{x_{17}} \text{ADAM} \urcorner$ and $\ulcorner \text{LO } x_{17} \text{ KELEV} \urcorner$ are scrambledly satisfied by s^{**} , so $\ulcorner_{x_{17}} \text{ADAM} \urcorner$ is scrambledly satisfied by s^{**} and $\ulcorner_{x_{17}} \text{KELEV} \urcorner$ is not scrambledly satisfied by s^{**} , so the image under μ^{-1} of the scrambled denotation of $\ulcorner_{x_{17}} \urcorner$ relative to s^{**} —which is just the 17th member of s^{**} —is human and not a dog, and so something in U is human and not a dog. This completes the demonstration that $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is satisfied by s iff $\ulcorner \text{MASHEHU } x_{17} x_{17} \text{ADAM VE LO } x_{17} \text{ KELEV} \urcorner$ is scrambledly satisfied by s .

References

- Dahllöf, Mats. (1999) 'Predicate-Functor Logic', *Reports from Uppsala University Department of Linguistics* 34, 101–123.
- Davidson, Donald. (1974) 'Belief and the Basis of Meaning', *Synthese* 27, 309–323.
- Davidson, Donald. (1977) 'Reality Without Reference', *dialectica* 31, 247–258.
- Devitt, Michael. (1983) 'Realism and the Renegade Putnam: A Critical Study of Meaning and the Moral Sciences', *Noûs* 17, 291–301.
- Donnellan, Keith. (1966) 'Reference and Definite Descriptions', *Philosophical Review* 75, 281–304.
- Hawthorne, John. (2007) 'Craziness and Metasemantics', *Philosophical Review* 166, 427–440.
- Heck, Richard. (2014) 'Semantics and Context-Dependence: Towards a Strawsonian Account', in Burgess, Alexis and Sherman, Brett (eds.), *Metasemantics: New Essays on the Foundations of Meaning* (Oxford: Oxford University Press).
- Hale, Bob and Wright, Crispin. (1997) 'Putnam's Model-Theoretic Argument against Metaphysical Realism', in Wright, Crispin and Hale, Bob (eds.), *A Companion to the Philosophy of Language* (Oxford: Blackwell).
- Kaplan, David. (1989) 'Demonstratives', in Almog, Joseph, Perry, John, and Wettstein, Howard (eds.), *Themes from Kaplan* (New York: Oxford University Press).
- King, Jeffrey. (2014) 'Speaker Intentions in Context', *Noûs* 48, 219–237.
- Kripke, Saul. (1980) *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Kripke, Saul. (1992) 'Logicism, Wittgenstein, and *De Re* Beliefs about Numbers', unpublished.
- Lewis, David. (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61, 343–377.
- Lewis, David. (1984) 'Putnam's Paradox', *Australasian Journal of Philosophy* 62, 221–236.
- Merrill, Gary H. (1980) 'The Model-Theoretic Argument against Realism', *Philosophy of Science* 47, 69–81.
- Millikan, Ruth Garrett. (1989) 'Biosemantics', *Journal of Philosophy* 86, 281–297.
- Millikan, Ruth Garrett. (1990) 'Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics', *Philosophical Topics* 18, 151–161.
- Putnam, Hilary. (1981) *Reason, Truth and History* (Cambridge: Cambridge University Press).
- Putnam, Hilary. (1990) *Realism with a Human Face* (Cambridge, MA: Harvard University Press).
- Quine, W.V. (1960) 'Variables Explained Away', *Proceedings of the American Philosophical Society* 104, 343–347.
- Quine, W. V. (1968) 'Ontological Relativity', *Journal of Philosophy* 65, 185–212.
- Schwartz, Wolfgang. (2014) 'Against Magnetism', *Australasian Journal of Philosophy*.
- Sider, Theodore. (2011) *Writing the Book of the World* (New York: Oxford University Press).
- Simchen, Ori. (2012) *Necessary Intentionality: A Study in the Metaphysics of Aboutness* (Oxford: Oxford University Press).
- Simchen, Ori. (2013) 'Token-Reflexivity', *Journal of Philosophy* 110, 173–193.
- Simchen, Ori. (2015) 'Metasemantics and Legal Interpretation', in Pavlakos, George and Rodriguez-Blanco, Veronica (eds.), *Reasons and Intentions in Law and Practical Agency* (Cambridge: Cambridge University Press).

- Simchen, Ori. (unpublished) *Semantics, Metasemantics, Aboutness*.
- Weatherson, Brian. (2012) 'The Role of Naturalness in Lewis's Theory of Meaning', *Journal for the History of Analytical Philosophy* 1, 1–19.
- Williams, J. Robert G. (2007) 'Eligibility and Inscrutability', *Philosophical Review* 116, 361–399.
- Wittgenstein, Ludwig. (1961) *Tractatus Logico-Philosophicus*, translated by Pears, David and McGuinness, Brian (London: Routledge and Kegan Paul).