
+ • +

THE JOURNAL OF PHILOSOPHY

VOLUME CX, NO. 4, APRIL 2013

+ • +

TOKEN-REFLEXIVITY*

Token-reflexivity is commonly understood as reference of a token to a token of which it is a part, proper or not. It may be usefully contrasted with a familiar formal kin: what is achieved with a certain singly universally quantified dyadic predicate of the language of formalized type theory predicated of the formal numeral of the number encoding that singly universally quantified dyadic predicate with a free variable y .¹ In what follows the possibility of the latter type of construction in a formal setting will provide a stark point of contrast with token-reflexivity understood as token self-reference, a purported species of natural phenomena, with the token-reflexives themselves understood as the bearers of self-reference. My main aim will be to show that there is no token-reflexivity thus understood, and so, no token-reflexives. The comparison with reflexive constructions à la Gödel will provide a background against which to discuss the centrality of conditions of production—as opposed to conditions of consumption—in the study of natural language.

We begin by distinguishing two relations that signs can bear to things signified: *conferred denotation* and *produced reference*. An expression bears the conferred-denotation relation to a thing solely by being interpreted as standing for the thing in question. The relation of conferred denotation is interpretive: a bears this relation to b by being interpreted as standing for b .² I overhear people at a party speaking of someone named ‘Bertie’ and surmise that they are speaking of Bertrand Russell. Right or wrong, my thus surmising establishes a conferred-denotation relation between the tokens of ‘Bertie’ I encounter

*This paper is dedicated to the memory of Ruth Barcan Marcus.

¹Throughout the paper I follow the usual practice of letting use/mention ambiguities be settled by the context.

²And the interpretive buck can stop here: for a to be interpreted as standing for b need not require that a be interpreted as being interpreted as standing for b .

and Russell. An analogy: I walk into a room with a chessboard in middle game where the piece for the black queen is missing. In an effort to make sense of the game I regard a bottle cap occupying one of the squares as standing for the black queen. The bottle cap interpreted as standing for the black queen is akin to the conferred-denotation relation that the encountered tokens of 'Bertie' bear to Russell.

The relation of produced reference, on the other hand, is not an interpretive relation. a bears the produced-reference relation to b by being employed to stand for b . One way to think about what it takes for someone to employ a to stand for b is to think of a as produced with an intention to refer to b . And we may think of such an intention as a *de re* cognitive attitude, a cognitive relation that the speaker bears to b in particular.³ The speakers at the party were really talking about Bertie Higgins by intending to refer to Higgins with their tokens of 'Bertie'. Matters of interpretation aside, their tokens of 'Bertie' bore the produced-reference relation to Higgins. By way of analogy again, you and I are about to play chess and notice that the piece for the black queen is missing. We resolve to use a bottle cap. Our employing the bottle cap for the black queen makes it so that the cap stands for the black queen. And this remains so whether an onlooker deems it so, or regards it, rather, as an incidental piece of trash that found its way to the board. In short, that a bears the produced-reference relation to b is determined by a 's conditions of production.⁴ On the other hand, that a bears the conferred-denotation relation to b is determined solely by a 's conditions of consumption.

Employing the distinction between conferred denotation and produced reference allows me to refine my main thesis: There is no token-reflexivity as produced self-reference. We begin our exploration of why this is so by turning to examine the classic example of conferred self-denotation found in the celebrated metamathematical work of Gödel. Gödel's construction of a sentence of the language of formalized type theory L_p that is neither provable nor refutable in the

³See section IV below. I make an extended case for this way of thinking of token production as the outcome of suitable referential (and phonological or graphological) intentions, and for thinking of referential intentions as attitudes directed at particular things, in chapter 3 of Ori Simchen, *Necessary Intentionality: A Study in the Metaphysics of Aboutness* (New York: Oxford, 2012). A referential intention, on this way of looking at things, is a real-world relation borne to a morphological item on the one hand and the item to be referred to on the other.

⁴Such production may involve the utilization of found objects—by analogy to producing the black queen for the game out of the bottle cap—as in the assembly within the ransom note of a referential 'your son' from letters cut out of a magazine.

system of formalized type theory P (on the condition of ω -consistency⁵) is widely and misleadingly regarded as self-referential.⁶ How such a characterization is misleading is something we will have to explore in some detail below. As will emerge, for a construction to be self-referential is for part of it to be employed to refer to the whole of which it is a part, a very different matter from regarding a syntactic construction as self-denotative. As regarding self-referentiality, one would expect such reflexivity to be evidenced by a construction's syntax, by the occurrence of a referring expression purportedly referring to the whole. But in the case of Gödel's original construction such evidence actually points in the wrong direction, as we will see. This, however, in no way prevents us consumers of the metamathematical result from regarding the construction as self-denotative.

I. CONFERRED SELF-DENOTATION: THE GÖDEL SENTENCE

In his classic 1931 paper Gödel says the following about the sentence of L_p that has come to be known as the Gödel sentence:

We therefore have before us a proposition that says about itself that it is not provable [in PM].⁷

Gödel then hastens to add a footnote:

Contrary to appearances, such a proposition involves no faulty circularity, for initially it [only] asserts that a certain well-defined formula (namely, the one obtained from the q^{th} formula in the lexicographic order by a certain substitution) is unprovable. Only subsequently (and so to speak by chance) does it turn out that this formula is precisely the one by which the proposition itself is expressed.⁸

The first passage, belonging to the heuristic first part of Gödel's paper, encourages the misleading impression of the achievement as having shown that a sentence of L_p a part of which refers to the

⁵ The system is ω -consistent if for no formula ϕ with a free variable, $\vdash \neg \forall x \phi(x)$ while for every number n , $\vdash \phi(n)$ (where n is a formal numeral—a sequence of n 'S's followed by '0' where 'S' is the sign for the successor function). ω -consistency implies consistency. I adapt the original formal language L_p to a more familiar notation to facilitate readability throughout.

⁶ For example, many instances of this tendency are peppered throughout chapter XVI of Douglas R. Hofstadter's highly influential *Gödel, Escher, Bach: An Eternal Golden Braid* (New York: Basic, 1979).

⁷ Kurt Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in Jean van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931* (Cambridge: Harvard, 1967), pp. 596–616, at p. 598. The English translation of Gödel's paper is due to van Heijenoort and approved by Gödel. The square brackets contain Gödel's own additions to van Heijenoort's translation.

⁸ *Ibid.*

sentence itself is neither provable nor refutable in the formal system (on the condition of ω -consistency). The second passage, while itself not without difficulty in allowing the intrusion of epistemic matters into the metaphysics of the situation—as illustrated by the phraseology of “initially” and “subsequently”—tries at least to correct the impression engendered by the first passage.

Let us be reminded of what it is in the metamathematical setting that encourages the impression of self-reference. We need not embark on a detailed examination of Gödel’s monumental results and their ramifications—a bare outline of the general strategy should suffice.

After laying out system P , Gödel devises an effective correlation of signs of L_P with numbers and then of strings of signs with numbers utilizing the Fundamental Theorem of Arithmetic. Through this correlation (“Gödel numbering”) syntactic properties such as being a formula and syntactic relations such as being a proof of a formula become associated with properties of and relations among numbers. Much of Gödel’s paper is devoted to showing that syntactically significant number-theoretic properties and relations, including the two just mentioned, are primitive recursive—that they are definable by means that are effectively decidable. And Gödel shows that primitive-recursive properties and relations are *numeralwise expressible* in the formal system P : for any such m -place relation R there is an m -place formula $\phi(x_1, \dots, x_m)$ such that $\vdash \phi(\mathbf{n}_1, \dots, \mathbf{n}_m)$ if R obtains with respect to n_1, \dots, n_m , and $\vdash \neg \phi(\mathbf{n}_1, \dots, \mathbf{n}_m)$ if R fails to obtain with respect to n_1, \dots, n_m (where for any number n , the formal numeral \mathbf{n} is a sequence of n ‘S’s followed by ‘0’). It turns out that the relation R_g that a number m bears to a number n if and only if it is not the case that m encodes a proof the last line of which is the sentence that results from the formula encoded by n when the free variable y in that formula is replaced by the formal numeral \mathbf{n} , is itself primitive recursive. And so, this relation R_g is numeralwise expressible. Letting $\phi_g(x, y)$ numeralwise express R_g , the Gödel sentence—the one that is neither provable nor refutable on the condition of ω -consistency—is just $\forall x \phi_g(x, \mathbf{n}_g)$, where \mathbf{n}_g encodes the formula $\forall x \phi_g(x, y)$.⁹

⁹ *Not provable if system is ω -consistent*: Assume that $\vdash \forall x \phi_g(x, \mathbf{n}_g)$. Then it has a proof: let m encode it. Given that \mathbf{n}_g encodes $\forall x \phi_g(x, y)$, by definition of R_g , R_g does not obtain with respect to m, \mathbf{n}_g , and so $\vdash \neg \phi_g(\mathbf{m}, \mathbf{n}_g)$. And yet from the initial assumption it follows that $\vdash \phi_g(\mathbf{m}, \mathbf{n}_g)$. So the system is inconsistent and therefore ω -inconsistent.

Not refutable if system is ω -consistent: Assume that the system is ω -consistent. Then by previous result $\nvdash \forall x \phi_g(x, \mathbf{n}_g)$. So every number n bears the relation R_g to the number encoding the result of replacing the variable y in the formula encoded by \mathbf{n}_g —that is, $\forall x \phi_g(x, y)$ —with the numeral \mathbf{n}_g . And so, for every number n , $\vdash \phi_g(\mathbf{n}, \mathbf{n}_g)$, so that by the system’s assumed ω -consistency $\nvdash \neg \forall x \phi_g(x, \mathbf{n}_g)$.

Where in all this is the reflexivity? We have before us a remarkable feat of conceptual engineering. In an effort to appreciate how it works we can take it apart and examine its crucial moving parts in isolation before putting it back together again.

Even from our very brief sketch of Gödel's achievement it should be clear that what we have here is not produced self-reference but at most conferred self-denotation. If $\forall x\phi_g(x, \mathbf{n}_g)$ says about itself, in any sense, that it is not provable, then this is a by-product of our regarding it in a certain way. However, it is not the syntax of the sentence itself that inclines us to regard it this way. The reflexivity effect is achieved as a combined effort (so to speak) of the predicate $\phi_g(x, y)$ and the numeral \mathbf{n}_g . The first numeralwise expresses the relation that holds between two numbers just in case it is not the case that the first number encodes a proof the last line of which is the result of replacing the free y in the formula that the second number encodes with the formal numeral whose number of 'S's encodes the formula. The second is the formal numeral whose number of 'S's encodes $\forall x\phi_g(x, y)$. Putting the two aspects together gets us the joint effect of reflexivity. But in no way does the syntax of $\forall x\phi_g(x, \mathbf{n}_g)$ attest to self-reference: the formal numeral \mathbf{n}_g is a sequence of 'S's whose length is the number encoding a *different* syntactic item—the formula $\forall x\phi_g(x, y)$ —followed by '0'. We may then interpret if we wish the entire sentence $\forall x\phi_g(x, \mathbf{n}_g)$ as saying about itself that it is not provable by recalling the number-theoretic relation that $\phi_g(x, y)$ numeralwise expresses.¹⁰ But nothing in the syntax of $\forall x\phi_g(x, \mathbf{n}_g)$ itself should be taken to suggest the bearing by anything of the produced-reference relation to $\forall x\phi_g(x, \mathbf{n}_g)$.

There is of course a shorter route to appreciating the absence of any produced reference here. The construction in question in no way depends on formal numerals standing for certain numbers, on the predicate $\phi_g(x, y)$ that numeralwise expresses the number-theoretic relation of interest standing for the relation in question, or on any other semantic matter. The achievement is purely syntactic.¹¹ As is familiar, we can prove a semantic version of Gödel's result—that there is a sentence of L_p that is true under the intended interpretation and is neither provable nor refutable in P under the

¹⁰ We may think of this as extending the notion of conferred denotation to whole sentences in a way that does not require of any particular constituent expression of a sentence to stand—by way of conferred denotation again—for what the sentence is interpreted as being about.

¹¹ This general point also tells against the idea that we have here a kind of self-satisfaction. I return to self-satisfaction in the discussion of descriptive token-reflexivity in section v.

assumption that P is sound with respect to the intended interpretation. But this appeal to interpretation is an obvious case of conferred denotation: the relation between signs and things signified is established solely by our regarding the signs as standing for the things in question. Under the auspices of conferred denotation, for a sentence to say anything about itself can be at most an instance of conferred self-denotation.

The point that reflexivity à la Gödel does not involve produced self-reference but at most conferred self-denotation is seen even more clearly when we consider the generalization known as the Diagonal Lemma or Fixed Point Theorem.¹² Here all syntactic details are washed away. The result states that for any formula $\phi(y)$ there is a sentence ψ such that $\vdash \phi(n) \leftrightarrow \psi$ where n encodes ψ . Any such ψ is a "fixed point" of $\phi(y)$.¹³ Lest it be thought that a fixed point of a predicate is a sentence whose syntactic structure somehow betrays that it says of itself that the predicate applies to it, we note that in general a predicate has multiple fixed points. And while the theorem offers a minimal characterization of the syntactic structure of the predicate as monadic, it makes no appeal whatsoever to the specific syntactic features of the predicate's fixed points. This, again, does not prevent us from describing a predicate's fixed point as saying of itself that the predicate applies to it as a matter of conferred self-denotation. But if we were to stick to the details of the result we would be at most warranted in describing an arbitrary fixed point of a predicate as an occupant of one side of a provable biconditional the other side of which is the predicate predicated of the formal numeral of the number encoding the initial side. And again, if we assume that saying *of* something that it is thus and so imports an appeal to semantics, and given that no such appeal is made in any of the above, there is nothing in the above that says of itself anything at all. In particular, nothing here says of itself that it is ϕ .

II. 'NEVER MIND'

Germinating out of mathematical logic, much of contemporary philosophy of language has not sufficiently absorbed some of the fundamental ways in which natural language differs from formal languages. The issue most pertinent to us here surrounds the existence of effective

¹²The result is due to Carnap. See §35 of Rudolf Carnap, *The Logical Syntax of Language* (London: Routledge and Kegan Paul, 1937).

¹³Letting $\phi(y)$ be the formula $\forall x \neg B(x, y)$, where $B(x, y)$ numeralwise expresses the number-theoretic relation B that a number encoding a proof bears to the number encoding the proof's last line, Gödel's first limitative result can follow with respect to a fixed point of this choice of formula.

procedures for determining whether or not a string of signs constitutes a formula in a formal language and whether or not a sequence of formulas constitutes a proof in a formal system. As witnessed above, the existence of such procedures was pivotal to the great metamathematical advances of the 1930s and their subsequent utilization. Only thus could such syntactic properties as being a formula in L_p or being a proof in P become associated with certain number-theoretic properties. And yet such features of formal languages and systems and their subsequent inspiration for the philosophy of language have tended to occlude the centrality of subsentential significance for natural language.

Speakers often produce linguistic expressions without ever getting a chance to complete their sentences. Sometimes others will not give them a chance to do so; other times they change their mind midway about what they wanted to say; and other times still they lose interest or just stop short in the interest of time. This range of phenomena is clearly distinct from another range that has received considerable attention in the literature surrounding pragmatic enrichment—the production of sentence fragments with whole sentential significance (for example, ‘Nice shirt!’ meaning roughly that the addressee is wearing a nice shirt). In the cases that interest us, on the other hand, speakers crucially stop short of completing their sentences and expressing what whole sentences would express. And yet for all that they produce items endowed with significance. So here it is not only the produced tokens but the significance itself that is subsentential. The phenomenology of subsentential significance suggests an intuitive test for specificational success, a test I will now call the Never Mind (or NM) test.¹⁴

A speaker says ‘President Obama... never mind’ and succeeds thereby in specifying Obama. This is evidenced by the obvious acceptability of the retort ‘Yes, what about him?’. A speaker says ‘President Obama told... never mind’ and succeeds thereby in specifying something requiring two kinds of completion. This is evidenced by the acceptability of the retort ‘Whom did he tell what?’ (or ‘What did he tell whom?’). Such verdicts of acceptability may be contrastive. We may consider which of several retorts would be more appropriate given the initial utterance. A speaker says ‘The person I met at the party last night... never mind’, and the appropriateness of the retort ‘What about *this* person?’ relative to ‘What about *such* a person?’ provides intuitive support for the speaker’s success in specifying a person rather than a property of persons.

¹⁴“Specificational” rather than “referential” in the interest of casting the net wider than just the cases of referential employment of noun phrases—see below.

Amenability to the NM test generally attests to the correctness of an incremental productivist metasemantics whereby speakers produce semantic significance in the temporal course of producing their tokens. This may be contrasted with a nonincremental alternative whereby significance emerges only with respect to sentence-long tokens. (Interpretationism is the most familiar form of nonincremental metasemantics, a point to which I return below.) But we should not expect too much of the NM test. Being an intuitive test for subsentential significance, it can easily deliver inconclusive results when it comes to what exactly the producer of the truncated utterance has managed to specify. A speaker says 'The first child to be born in the twenty-second century... never mind'.¹⁵ Informants appear to be divided over which retort would be more appropriate, 'What about *such* a child?' or 'What about *this* child?'. The indecision reflects a genuine lack of clarity over whether the speaker has succeeded in specifying a property of children or a particular child. We, as theorists, can step in and intervene at the service of considerations of systematicity in our semantics. Or we can ask, at the service of our overall metasemantics, how the speaker could have specified a particular child answering the condition of being firstborn in the twenty-second century if none such exists as of yet.

Cases of ambiguity, lexical or other, seem to show that sometimes we had better wait until the end of the utterances we encounter in order to assess what was said thereby. This, however, has little bearing on the current issue. A speaker says 'The bank... never mind', to which an interlocutor may well respond with 'Yes, what about it?'. The intuitive acceptability of such an exchange testifies to the fact that the speaker succeeded in specifying something definite, a financial institution or a side of a river as the case may be, notwithstanding an interpreter's potential ignorance as to what kind of thing is at issue.

Discourses surrounding truncated utterances as illustrated above suggest (a) that speakers can in fact succeed in specifying something definite by their truncated utterances, and (b) what in particular they manage to thus specify. As we saw with the example of 'the first child to be born in the twenty-second century', we must exercise caution when it comes to (b). And in any case, what speakers manage to specify in the course of token production is subject to theoretical constraints from our overall semantics and metasemantics. But the NM test offers strong intuitive support for (a). I submit that amenability to the NM test can be counted on to show that an expression

¹⁵The example is from section IX of David Kaplan, "Quantifying In," *Synthese*, XIX, 1/2 (December 1969): 178–214.

succeeds in specifying something, but cannot be counted on to deliver a precise rendition of what the expression succeeds in specifying in the course of token production.

On the consumption side of linguistic exchange psycholinguistic evidence suggests that audiences interpret tokens incrementally in the temporal course of being presented with subsentential expressions rather than all at once in the presence of finished sentence-long products.¹⁶ The verdicts of the NM test suggest a corollary on the side of production: speakers produce significance incrementally in the temporal course of producing a string of subsentential expressions rather than all at once with respect to finished sentence-long products. Because the voluntary cessation of speech or its interruption by another can occur arbitrarily close to the time of completion of production of the token that makes an incremental contribution to overall significance, we can formulate the following principle of simultaneity of production of tokens and their incremental contribution to overall significance:

(SPC) For any subsentential token *T* whose significance is added incrementally to the significance of a whole in which it partakes, *T* makes its contribution to overall significance as soon as *T* is produced.¹⁷

We note that token production under (SPC) is to be taken to include the utilization of a preexisting item for semantic purposes, such as the transfer of a certain amount of chalk from the chalk piece in one's hand to the blackboard, but that details surrounding the metaphysics of token production are subtle and not very well understood. An important implication of (SPC) for present purposes is that significant tokens are produced *as* significant, rather than being produced first and endowed with significance later.¹⁸ This, once again, is strongly

¹⁶Such evidence most notably includes garden-path sentences—for example, 'The horse raced past the barn fell' and 'The old man the ship'—where in parsing the material we are initially led down the wrong path.

¹⁷We emphasize that the principle governs token production rather than token consumption. Even for cases of cataphora such as 'She was hesitant at first, but after a while Jane decided to leave', where from the point of view of consuming the token the interpretation of 'she' is beholden to that of the later occurring 'Jane', from the point of view of production there is no such dependence. (If upon beginning with 'She...' the speaker were rudely interrupted by 'Yes, yes, what about her?', rudeness aside, the *prima facie* acceptability of the interruption suggests that the cataphoric 'she' succeeds in making its incremental contribution to overall significance at the time of its production, matters of interpretation aside.)

¹⁸A natural theoretical counterpart to the incremental metasemantic view articulated here within the syntax/semantics interface is work that pursues a surface-compositional semantics (under the hypothesis of locality of interpretation) within a Categorical Grammar framework. See Pauline Jacobson, "Towards a Variable-Free

suggested by the possibility of an arbitrarily close temporal proximity between cessation of speech and amenability to the NM test. (SPC) will prove crucial for the discussion of token-reflexivity to follow.

III. PRODUCTIVISM AND INTERPRETATIONISM

The question of the existence of token-reflexivity as produced self-reference demands a discussion of the prior and very general metasemantic question of what gives rise to semantic significance. In this large theoretical context, produced reference and conferred denotation belong to two very different orientations for answering the metasemantic question—productivism and interpretationism.

From a productivist standpoint the metasemantic question primarily targets the conditions of producing an item of significance. The basic idea here is that whatever the conditions of consumption of the expressive product may be, there are facts surrounding the item's conditions of production that have to be in place in order for the product to have the significance it has. From an interpretationist standpoint, on the other hand, the metasemantic question only targets conditions of consumption. To be endowed with significance, on such a view, is to be interpreted as such.¹⁹ On the latter way of looking at things the achievement of such endowment, if it can be put this way, is on the consumer side rather than on the producer side of linguistic exchange.

To appreciate the contrast more fully it is useful to consider an analogy in the case of a humdrum artifact, a particular hammer. Let us say that the hammer has a certain function or purpose: to drive in nails. To the metateleological question of how it is that this particular item came to have this particular purpose a metateleological productivist would answer that having such a purpose is determined by the item's conditions of production, conditions that plausibly include the intention to produce an item for the purpose in question.

Semantics," *Linguistics and Philosophy*, xxii, 2 (April 1999): 117–84. See also Chris Barker and Jacobson, "Introduction: Direct Compositionality," in Barker and Jacobson, eds., *Direct Compositionality* (New York: Oxford, 2007), pp. 1–19. For recent work within this research program see Mark Steedman, *Taking Scope: The Natural Semantics of Quantifiers* (Cambridge: MIT, 2012).

¹⁹ Or be interpretable. Interpretationists may not insist on the existence of actual linguistic exchange to generate significance, resorting instead to a dispositional account. And yet even such a shift in focus from actual interpretation to interpretability does not detract from the general point that significance on such an account emerges from conditions of consuming linguistic expressions rather than producing them. For further discussion, see Alex Byrne, "Interpretivism," *European Review of Philosophy*, III (1998): 199–223.

A metateleological interpretationist, on the other hand, would view endowment with such a purpose as determined only by how the item is regarded. As against the productivist, the interpretationist might point out that the particular hammer under consideration was mass-produced and not the product of any individual intention vis-à-vis this very item. As against the interpretationist, the productivist might adduce a fundamental distinction between items created for the purpose of driving in nails and items whose features make them only accidentally suited for such a task—found rocks, say—arguing that the interpretationist smooths over such important differences.

Once we see our explananda as a species of natural phenomena it is difficult not to view them under the auspices of some version or other of productivism.²⁰ Things are generally the way they are due to how they came to be. It seems perniciously revisionary to suppose that the various ways the relevant phenomena are interpreted confer all the relevant characteristics unto them regardless of how they were in fact produced. Semantic facts with respect to tokens, or teleological facts with respect to artifacts, may indeed be essentially relational vis-à-vis consumers, themselves denizens of the natural order. It may indeed be the case that in order for there to be a hammer in the world there has to be a characteristic purpose for such a thing, and for this to come about there has to be an audience for such a thing so that it is strictly speaking impossible to create the item and sustain its purpose without regarding it in a certain way. Arguably, when a Paleolithic hominid created a hammer, the hammer had to be regarded as a tool with its characteristic purposes, at least by its creator. And yet such an item would not come into existence without certain conditions of production being in place, plausibly intentions on the side of its creator, including perhaps the intention that the item be regarded in a certain way. Of course hammers today are typically mass-produced, as are printed tokens of words. But it is no less true of such items that they are products of intentions. (And in this vein we should be thinking of mass-production not only as production *of* a mass—a mass of hammers or printed tokens as the case may be—but also as production *by* a mass—a mass of producers, a team.)

The topic surely deserves a more thorough treatment than I can offer it here. But having gone through this brief detour on the two general metasemantic orientations, productivism and interpretationism, we are finally ready to confront the issue of token-reflexivity as produced

²⁰I make an extended case for productivism and its ramifications in *Necessary Intentionality*.

self-reference head on. Reflection on the conditions of production of purported instances of the phenomenon will reveal it as spurious.

IV. PRODUCED SELF-REFERENCE: RTR

We can distinguish two alternatives for produced, rather than conferred, token-reflexivity: the standard referential variety of token-reflexivity RTR (for example, a token of 'This is ψ ' where the token of the demonstrative pronoun is supposed to refer to the entire sentence token) and a descriptive variant DTR (for example, a token of 'The token occupying region R is ψ ' occupying region R).²¹ Our main focus will be on the former. In the next section we will consider the latter alternative as well.

By way of introducing the set of issues relevant for RTR let us examine the example of DTR just given: a token of 'The token occupying region R is ψ ' occupying region R , where the description 'the token occupying region R ' is not being used referentially. To achieve the reflexivity effect it is crucial that the token in question should really occupy region R , where ' R ' names the very region occupied by the token. But consider the following: how is it that the name ' R ' succeeds in naming the very region occupied by the token? The question has a funny ring to it—presumably we could just *stipulate* ' R ' to be the name of the region occupied by the token and thereby achieve the reflexivity effect. What could be easier? But on second thought, we need to consider *when* this name could have been introduced. Presumably, it could not have been introduced *after* the token in question was produced if the token was to succeed somehow in specifying itself. For in that case the occurrence of ' R ' within the relevant token of 'The token occupying region R is ψ ' would be an occurrence of an empty name, which would prevent the successful expression of self-specification.²² Nor could the introduction happen *during* the production of the token. For it seems gratuitous to suppose that just as I was producing the token I was also stipulating a use for ' R ' without actually tokening the stipulation.

Could ' R ' have been introduced *before* the production of the token? Yes, but not in the way one might expect. Suppose I said: 'I hereby name the region occupied by the next sentence token I shall produce " R ". Have I succeeded in naming anything? It is hard to see how I could have been successful. The means by which the region is demarcated—the produced token—does not exist at

²¹ Within this typology we group referential descriptive token-reflexivity with RTR rather than with DTR.

²² Recall that we are assuming here that the token of 'the token occupying region R ' is not being used referentially. See previous footnote.

the time of the stipulation. Suppose I drop dead before managing to produce another token. In that case my stipulation for '*R*' would surely misfire. So for the stipulation to go through the future has to go on in a certain way: I have to produce a second sentence token. This means that before the second sentence token is produced the stipulation has not yet been completed. And so, before the second sentence token is produced the name '*R*' does not yet name anything. And so, the naming ceremony we are considering is not a real option for securing the reflexivity effect for the second sentence token—in the actual course of producing the sentence token, the token of '*R*' would fail to refer. The remaining option for pre-sentence-token stipulation is to name a region without reference to a future entity. I can say 'I hereby name the first blank page of my copy of Prior's *Objects of Thought* "*R*". I can certainly pull off such a naming ceremony. I can then open my copy of the book to *R* and scribble a token of 'The token occupying region *R* is ψ '.

The reason we could not introduce the name '*R*' by referring to a token of 'The token occupying region *R* is ψ ' was that at the time of the introduction the sentence token would not have existed. Similarly, I cannot introduce the name 'Johnny' with the decree 'I hereby name the next sentence token I shall produce "Johnny"', and then, hoping to achieve self-reference, issue a token of 'Johnny is ψ '. For in order that the stipulation go through, the future has to go on in a certain way. And so, before the second sentence token is produced the stipulation has not gone through yet. And for reasons analogous to those given above, such a naming ceremony at the service of achieving reflexivity can occur neither during the production of the token of 'Johnny is ψ ' nor after its production.

Our lesson is general: reference of the sort achieved by referring tokens of names, referring tokens of demonstrative pronouns, but also referring tokens of referentially used descriptions, requires the referent's antecedent existence. (Henceforth we reserve "reference" for the relevant sort and drop the qualification.) For me to name a ship by declaring 'I hereby name this ship "Queen Elizabeth"' and smashing a bottle against its side requires that the ship I purport to name exist at some time or other before it was referred to with 'this ship'.²³ And the same can be said if I declared instead 'I hereby name the boat "Queen Elizabeth"' about my referential use of 'the

²³ I assume that complex demonstratives are referential rather than quantificational, but not much turns on this for present purposes. Proponents of the quantificational approach may substitute a referential demonstrative pronoun for the complex demonstrative without loss.

boat'. (We imagine it is really a cruise ship and not a ferryboat I am naming.) Incidentally, reference to a thing does not seem to require the thing's continued existence into the moment of its being named. We can look up at the night sky and resolve to name one of the stars we see there 'Alpha', not realizing that the source of the radiation we were subjected to is a star that no longer exists. So we managed to name a thing of the remote past. We can do this because we lie in that star's causal wake, whether or not it presently exists. We can do this because being affected by portions of our surroundings takes (space)time.²⁴

Elsewhere I defend a certain view of the conditions of production of referential tokens inspired partly by Bromberger and Halle's ontology of phonology and partly by Kaplan's metaphysics of words.²⁵ On the view in question, the relation of token reference, *Refer*, is the product of two distinct relations, the relation *Produce-by* that the token bears to a referential intention, and the relation *Specify* that the referential intention bears to a thing. For token *T* to refer to *o*, *T* has to bear the *Produce-by* relation to a particular referential intention *ri*, and *ri* has to bear the *Specify* relation to *o*. The view also includes a particular understanding of referential intentions. Within a typology of specific (a.k.a. "*de re*") and generic (a.k.a. "*de dicto*") cognitive attitudes, referential intentions fall under the former type. They are specific attitudes directed at particular things, much like believing Orcutt to be a spy and wanting a particular sloop are attitudes directed at particular things—Orcutt and the sloop in question, respectively. On this account, for every specific (or *de re*) cognitive attitude a distinction is drawn between the causal-historical connection ("c-relation") that must obtain between the agent of the attitude and the attitude's subject matter on the one hand, and the attitudinal relation itself on the other. A relevant instance of the c-relation is a cognitive prerequisite for the obtaining of the further attitudinal relation of, say, believing of Orcutt that he is a spy or wanting a particular sloop. As cognitive prerequisite, the relevant instance of the c-relation is not a higher-level cognitive task in its own right but a causal-historical precondition for specific (or *de re*) cognition. Now, a referential intention is an intention to employ a

²⁴ I realize that presentists would remain unmoved, but it is difficult to see how presentism would be reconciled with a productivist metasemantics in any case. The issue deserves a separate discussion elsewhere.

²⁵ See chapter 3 of Simchen, *Necessary Intentionality*; Sylvain Bromberger and M. Halle, "The Ontology of Phonology (Revised)," in Noel Burton-Roberts, Philip Carr, and Gerard J. Docherty, eds., *Phonological Knowledge: Conceptual and Empirical Issues* (New York: Oxford, 2000), pp. 19–37; and Kaplan, "Words," *Aristotelian Society Supplementary Volume*, LXIV (1990): 93–119.

morpheme for a particular thing. It is thus construed as a complex relation among agent, morpheme, item to be referred to, and a dyadic relation *mRefer* that is to obtain between the morpheme and the item to be referred to. In such a case, two instances of the c-relation are required: a causal-historical connection between agent and morpheme and a causal-historical connection between agent and item to be referred to.²⁶

The details of such an account of token reference are obviously controversial. However, the items of controversy are largely beside the point of our present concern, which is token-reflexivity of the referential variety RTR. On any plausible productivist metasemantics the use of a term to refer to a particular item requires at least having some causal-historical rapport with the item being referred to.²⁷ Assuming there to be no backward or simultaneous causation, for a token to refer to a thing at a certain time requires that the thing in question exist at an earlier time. We formulate the antecedent availability of the referent principle as follows:

(AAR) If token *T* refers to item *o* at time *t*, then for some δ , $\delta > 0$,
o exists at $t - \delta$.

This much is common ground for any plausible productivist metasemantics. And it is this minimal productivist core, together with the earlier metasemantic principle (SPC) and the relatively uncontroversial semantic thesis that a referring token contributes its referent to overall significance, which spells trouble for produced self-reference.

For consider the production of a token of '*a* is ψ ' where the token of *a* refers to thing *o*. Let us assume that the production of the token of *a* is complete at time *t* and for any ϵ , $\epsilon > 0$, the production of this token of *a* is still in progress at $t - \epsilon$.²⁸ Referring tokens have their

²⁶ I schematically represent this as:

$\langle C(\text{agent, morpheme}), C(\text{agent, object}) \rangle \textcircled{C} \text{RI}(\text{agent, } \langle \text{morpheme, object} \rangle, \text{mRefer}(_1, _2))$,

where 'C' stands for the c-relation and '©' stands for a non-truth-functional cognitive prerequisite connective. See chapters 3 and 5 of Simchen, *Necessary Intentionality*, for various refinements and replies to objections.

²⁷ For present purposes we sidestep controversies surrounding reference to particular abstracta.

²⁸ Whichever time unit we select, we set aside problems that might arise from letting ϵ take on values that are deemed too large. For some such values it might be protested that it is perverse to suppose that the production of the token of *a* was already "in progress" at $t - \epsilon$ (during the Triassic Period, say). For other such values it might be protested that there are compelling cosmological reasons for supposing that there are no such times as $t - \epsilon$ (because the value of ϵ places $t - \epsilon$ before the Big Bang).

significance added incrementally to the significance of wholes in which they partake, so by (SPC), the token of *a* makes its contribution to the overall significance of the token of '*a* is ψ ' at *t*. By the semantic thesis that referring tokens contribute their referents to overall significance, it is *o* that is thus specified by the token of *a* at *t*. And by (AAR), *o* exists at $t - \delta$ for some $\delta, \delta > 0$. Now assume for *reductio* that the token of *a* refers to itself and that t^* is the time at which the production of this token of *a* is complete and for any $\varepsilon, \varepsilon > 0$, the production of this token of *a* is still in progress at $t^* - \varepsilon$. By (SPC) and the semantic thesis, the relevant token of *a* specifies this very token of *a* at t^* . So by (AAR), this token of *a* exists at $t^* - \delta$ for some $\delta, \delta > 0$. But this contradicts our choice of t^* : simply instantiate ε with the instantiation of δ . We thus conclude that the token of *a* in question does not refer to itself after all. By similar reasoning, no token of *a* refers to *any* token of which it is a part. And so we conclude that there really is no RTR. *A fortiori*, there are no token-reflexives in the sense of tokens that are produced to refer to wholes of which they are parts, proper or not.

The argument is an argument from a productivist metasemantics to the nonexistence of token-reflexivity in the sense of produced self-reference. Aside from the metasemantic premises, the argument has a semantic premise as well—that referring expressions contribute their referents to the significance of wholes in which they partake—but this premise is relatively uncontroversial. If the argument is sound, it has a hitherto unnoticed implication for Reichenbach's proposed semantic equivalence of sentences of the form '*... ID ...*', where *ID* is a pure indexical or true demonstrative, and sentences of the form '*... [the $x: x$ is a ϕ of this token] ... x ...*', where a token of 'this token' is to stand for the whole sentence token of which it is a part.²⁹ Whatever else may be said for or against Reichenbach's theory as a semantic proposal, a productivist metasemantics is not a live option for any account that proposes to utilize it, such as the token-reflexive B-theoretic analysis of tense.³⁰ Also, the foregoing, if sound, has implications for the existence of reflexive thought, as in Parfit's

²⁹ For the classic semantic utilization of 'this token', see §50 of Hans Reichenbach, *Elements of Symbolic Logic* (New York: The Free Press, 1947). See also the influential semantic critique of this theory in section x of Kaplan, "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in Joseph Almog, John Perry, and Howard Wettstein, eds., *Themes from Kaplan* (New York: Oxford, 1989), pp. 481–563.

³⁰ See, for example, J. J. C. Smart, "The Tenseless Theory of Time," in Theodore Sider, John Hawthorne, and Dean W. Zimmerman, eds., *Contemporary Debates in Metaphysics* (Malden, MA: Blackwell, 2008), pp. 226–38.

example by way of Harman of the insomniac thinking ‘I am not going to fall asleep because of my having this very thought’.³¹ It might be that it is the very thinking of such words that is keeping the insomniac awake. But inasmuch as the case cannot be a case of produced self-reference but at most a case of conferred self-denotation, it involves the further task of interpretation. Given this additional cognitive burden, small wonder that the insomniac is kept awake.

V. PRODUCED SELF-SATISFACTION: DTR

But is there no token-reflexivity other than conferred self-denotation? Not if the foregoing argument is sound and the reflexivity effect is to be achieved via referential uses of expressions. Let us turn, however, to the descriptive variety DTR. We concentrate on the schema for DTR we have already encountered: a token of ‘The token occupying region R is ψ ’ occupying region R , where the token of ‘the token occupying region R ’ is not being used referentially. We consider three alternative construals: DTR along orthodox Russellian lines (O-DTR), along neo-Russellian lines (N-DTR), and along Strawsonian lines (S-DTR).

For the discussion of O-DTR we regard the Russellian theory applied to operator-free cases as the reduction of ‘The ϕ is ψ ’ to $\exists x(\forall y(\phi y \leftrightarrow y = x) \wedge \psi x)$. A token of ‘The token occupying region R is ψ ’ occupying region R would be deemed semantically equivalent to a token of $\exists x(\forall y(\text{token-occupying-}R^*y \leftrightarrow y = x) \wedge \psi x)$ occupying region R^* .³² Moreover, a general feature of such an analysis is that an apparently subject-predicate structure turns out to be an existentialization of a complex predicate in much the way that on the standard FOL analysis the apparently plural subject-predicate ‘All whales are mammals’ becomes a universalization of a complex predicate. Given the incremental production of significance, orthodox Russellianism, with its syntactic revisionism towards descriptive sentences, seems implausible in any case, matters of reflexivity aside. The orthodox Russellian analysis precludes specificational prospects for ‘the ϕ ’ in just the way that the FOL account precludes such prospects for ‘all whales’. In both cases

³¹ See Gilbert Harman, “Self-Reflexive Thoughts,” *Philosophical Issues*, xvi, 1 (September 2006): 334–45.

³² $R^* \neq R$: I am assuming here that the description in the analysandum is proper and that an adequate semantic analysis of the R -located token of ‘The token occupying region R is ψ ’ would be required to preserve the reflexivity effect. For further discussion of such a requirement, see Tyler Burge, “Self-Reference and Translation,” in F. Guenther and M. Guenther-Reutter, eds., *Meaning and Translation: Philosophical and Linguistic Approaches* (London, UK: Duckworth, 1978), pp. 137–53.

it is denied that the phrase contributes its own distinctive significance to the significance of the whole in which it partakes. And yet we find ‘What about *it/her/him?*’, or ‘What about *such* a thing/person?’, to be acceptable responses to ‘The ϕ ... never mind’, just as much as we find ‘What about *them?*’, and perhaps ‘What about *such* things?’, to be acceptable responses to ‘All whales... never mind’.³³ Orthodox Russellianism thus has an unfortunate metasemantic implication: it conflicts with the incremental generation of overall significance attested to so vividly by the amenability to the NM test. And so, I claim, O-DTR is not an option.

Russellianism can be modified in such a way as to be amenable to the NM test. On a neo-Russellian view, ‘The ϕ is ψ ’ has the structure of a noun phrase followed by a verb phrase. The truth conditions are those given by Russell—‘The ϕ is ψ ’ is true if and only if one and only one thing ϕ s and that thing ψ s—but the contextual elimination of the description is relaxed. According to such an account, in saying ‘The next lottery winner... never mind’ the speaker can succeed in specifying a function that maps any function f that maps individuals onto truth-values onto TRUE just in case one and only one individual is a next lottery winner and f maps this individual onto TRUE. These details are obviously not available to intuition; and yet, I claim, the incremental contribution to overall significance of the phrase ‘the ϕ is thus available.

Going back to the case that interests us, a token of ‘The token occupying region R is ψ ’ occupying region R is understood under the auspices of N-DTR as semantically equivalent to a token of

[the x : token-occupying-region- R^*x] ψx

occupying region R^* .³⁴ The quantified noun phrase specifies a function that maps the function denoted by ψ onto TRUE just in case one and only one individual is a token occupying region R^* and the function denoted by ψ maps it onto TRUE. We note that the token of the N-DTR analysis itself is nowhere being referred to in this analysis. By uniquely occupying R^* , the token of the N-DTR

³³We note again that when it comes to *what* the description contributes to overall significance the NM test need not be thought of as reliable. Surely this is the case from the point of view of current semantic theory whereby the contribution of ‘all whales’ to overall significance is that of a function mapping functions that map individuals onto truth-values onto truth-values. But the delivery of the NM test when it comes to the fact *that* the quantified noun phrase makes an incremental contribution to overall significance seems intuitively robust enough to be respected by any plausible metasemantics.

³⁴See footnote 32.

analysis satisfies the embedded condition of being a unique occupant of R^* , thus making the truth-value of the whole solely dependent on whether or not the function denoted by ψ maps that token onto TRUE. Instead of produced self-reference, which we cannot have, what we have here is produced self-satisfaction.

Moving on to S-DTR, the situation here is slightly more complicated. On Strawson's view the principal error in Russell's theory is the conflation of meaningfulness for descriptive sentences with such sentences expressing propositions on particular occasions of use.³⁵ So, in particular, the Russellian idea of (an operator-free) 'The ϕ is ψ ' logically implying there being a unique ϕ is mistaken. Rather, we are told, the *use* of such a sentence presupposes—rather than asserts—there being a unique ϕ . And if per chance there is no unique ϕ , then *pace* Russell we do not have falsity here but failure to express a proposition. And such failure is perfectly compatible with the meaningfulness of the sentence employed.

A case discussed by Strawson is that of a speaker saying 'The king of France is wise' during the reign of Louis XV. To ponder what the speaker said by considering the conditions of production of the token we need to go beyond Strawson's discussion and fill in further details. After Donnellan we might say that the speaker could have used the description referentially as one for Louis XIV, saying something whose truth or falsity depended on whether or not Louis XIV was wise at the time of utterance.³⁶ But setting aside such a referential use, suppose that the speaker intended to speak of whoever was in charge while mistakenly thinking it was Louis XIV. The speaker had just returned to France after years in exile and had not yet learned that Louis XIV had died and was succeeded to the throne by his great-grandson Louis XV. In intending to speak of whoever was in charge—wishing perhaps to seem faithful to the throne after all these years—the speaker had spoken falsely of the five-year-old successor to the throne. According to Strawson, the speaker would have simply employed the received means of speaking of the occupant of the highest office in the land.

We note in passing that the Strawsonian treatment of the case cannot be assimilated to a Donnellanian attributive use. In Donnellan's attributive cases, as in any version of the Russellian view, it is presumed

³⁵ See P. F. Strawson, "On Referring," *Mind*, LIX, 235 (July 1950): 320–44.

³⁶ For the classic discussion of such uses of descriptions, see Keith S. Donnellan, "Reference and Definite Descriptions," *Philosophical Review*, LXXV, 3 (July 1966): 281–304.

that the attribute of being king of France enters into the semantic computation of whether or not what is said is true. Not so here. For the Strawsonian, what enters the semantic computation is whether or not a certain individual—Louis XV, as it happens—is wise. (Perhaps we may think of such a case as involving a silent occurrence of Kaplan's 'dthat' under the operator interpretation.³⁷) The case is meant to be singular with respect to whoever happens to be the king.³⁸

But when it comes to the conditions of production of a token of 'The token occupying region R is ψ ' occupying region R and the question of reflexivity, the situation according to S-DTR is not so very different from the situation according to N-DTR. To claim that this token is ψ the speaker presupposes that one and only one token satisfies 'token-occupying-region- R '. So regardless of whether or not the condition of being such a token contributes to the semantic computation, the sentence token is about itself only to the extent that it (presupposedly) satisfies the condition uniquely. So under S-DTR we can have produced self-satisfaction as well, but with "satisfaction" understood to extend beyond the confines of semantic computation.

VI. CONCLUDING REMARKS

Could we not just introduce by fiat a token-reflexive pure indexical governed by the character that yields the token itself in context? Let tr be the type governed by the following character: for any context c , the denotation of tr in c is the token, if any, of tr in c . We could certainly introduce such an expression type and thereby achieve stipulated self-denotation. But tokens of tr would not be referential in the relevant sense, the sense in which referring tokens of names, referring tokens of demonstrative pronouns, and referring tokens of referentially used descriptions are referential.

If I am right, there is no token-reflexivity as produced self-reference. This does not mean, of course, that we are somehow barred from

³⁷ See Kaplan, "Dthat," in Peter Cole, ed., *Syntax and Semantics, Volume 9: Pragmatics* (New York: Academic, 1978), pp. 221–53. For a discussion of the operator versus demonstrative surrogate interpretation of 'dthat', see section 1 of Kaplan, "Afterthoughts," in Almog, Perry, and Wettstein, eds., *op. cit.*, pp. 565–614. The idea of a silent occurrence of 'dthat' is inspired by Strawson's own appeal to the example of the first-person singular pronoun in explaining how meaningfulness is related to referential employment on a particular occasion of use. See Strawson, *op. cit.*, pp. 326–27.

³⁸ We see this by comparing falsity conditions. For the Russellian, 'The king of France is wise' is false if and only if it is not the case that something or other is a unique king of France and wise. For the Strawsonian, on the other hand, the use of the sentence 'The king of France is wise' is false if and only if whomever the relevant use of 'the king of France' specifies is not wise.

regarding a particular syntactic construction as denoting itself. Sometimes regarding a piece of syntax as self-denotative has clear heuristic value, as anyone who has tried to teach limitative metamathematical results to novices can attest. Other times features of a token's syntax, combined with certain perceived aspects of the token's circumstances, incline us to regard the token as self-denotative, as in the case of an encountered, relatively isolated occurrence of an instance of 'This token is ϕ '. Indeed, on the production side one can resolve to produce a token with the aim that the item thus produced be interpreted as self-denotative. This would not be, however, the production of a self-referential token. A self-referential token would have to be a referential token, and, if I am right, referential tokens are both inevitably produced *as* referential and require the antecedent availability of their referents. In the case of self-reference this would mean that the token would have to exist before it existed. But nothing exists before coming into existence.

ORI SIMCHEN

University of British Columbia