Intensionality from self-reference

T. Parent (Nazarbayev University)
nontology@gmail.com

Abstract

This paper identifies a novel type of intensionality, which results from a self-referential operation defined only in formal terms. Since intentional operations are already excluded on principle, nothing here forces revisions to classical logic. But the paper reveals an unexpected route by which intensionality can arise, showing that additional vigilance is needed to prevent such constructions.

1. Introduction

Self-reference is prominent in various paradoxes and limitation theorems as from Gödel (1931) and Tarski (1933). But self-reference has been given surprisingly little attention on its own. The explanation seems to be that self-reference per se is seen as quite straightforward. Kripke (1975) exemplifies this attitude in a brief remark before moving on to other concerns:

> Let 'Jack' be a name of the sentence 'Jack is short', and we have a sentence that says of itself that it is short. I can see nothing wrong with "direct" self-reference of this type. If 'Jack' is not already a name in the language, why can we not introduce it as a name of any entity we please? In particular, why can it not be a name of the (uninterpreted) finite sequence of marks 'Jack is short'? (Would it be permissible to call this sequence of marks "Harry," but not "Jack"? Surely prohibitions on naming are arbitrary here.) There is no vicious circle in our procedure, since we need not interpret the sequence of marks 'Jack is short' before we name it. Yet if we name it "Jack," it at once becomes meaningful and true. (p. 693).

Undoubtedly, the 'Jack' example is innocuous—and more broadly, it is beyond question that a sentence can unproblematically index itself, as in the relevant limitation theorems. Conversely, it is well-known that self-reference in combination with semantic terminology (such as 'true' and 'denotes') yeilds pathology. But in a language free of semantic terms, is everything harmless that might be called "self-referential"? One might grant that a self-referential sentence like Jack is legit, yet still raise questions about self-referential *terms*, e.g., 'Jim', which denotes its own quotation. (It apparently follows that Jim = the quotation of

Jim = the quotation of the quotation of Jim...) There are further self-referring terms whose cogency might seem doubtful, such as 'the term identical to this very term', or 'this very quotation', or 'the term distinct from x that has just replaced 'x' in this very open term'.

Indeed, something like this latest example expresses an intensional operation, as we shall see. The main thesis, then, is that at least one kind of self-referential term creates intensional inconsistency, even in a language with neither semantic terminology nor the usual intensional elements (propositional-attitude verbs, modal operators, attribute abstraction, idioms like 'so called', etc.).[1]

Given that self-reference is prominent in classical logic, does this signal some kind of revolution?[2] Hardly. In fact, the following has no revisionary consequences at all. Again, the relevant inconsistency is due to an intensional operation, and such operations are already excluded on principle from classical logic. Nonetheless, the paper reveals an unexpected route by which intensionality can arise, showing that additional vigilance is needed to prevent such constructions.

2. Quotational intensionality refashioned

Consider a theory that includes denumerably many constants 'a_0', 'a_1', 'a_2'..., interpreted as follows (where $\underline{n}$ is the numeral for n):

(i) 'a_0' denotes 0

(ii) For all $n > 0$, $\ulcorner \mathrm{a}_{\underline{n}} \urcorner$ denotes $n - 1$

(Notation: Per Quine 1951, an expression with corner quotes denotes the concatenation of symbols enclosed in the corners, after the replacement of any metavariables. Thus, as concerns $\ulcorner \mathrm{a}_{\underline{n}} \urcorner$, if '$\underline{n}$' is replaced with '1', the corner-quoted expression then denotes the

[1] On such cases of intensionality, see Quine (1960, sections 30ff.).

[2] The term 'classical logic' is a bit of a misnomer. As stated in the preface to the first edition of Priest (2008), it was not the logic that was prominent in the classical era of Greece and Rome. It is thus perhaps better to call it "standard logic," as the logic that is typically taught in introductory symbolic logic courses. Yet let that pass; "classical logic" has become the convention regardless.

symbol 'a' concatenated with a subscripted '1'. I.e., it denotes 'a_1'.) Our theory establishes co-reference between 'a_0' and 'a_1', but as is well-known, the substitution of co-referring terms within quotation marks is invalid. Thus, from the truism that 'a_0' = 'a_0', substituting into quotation would allow us to derive the absurdity that 'a_0' = 'a_1'. Such a maneuver is therefore not permitted in classical logic.

What is much the same, a substitutable variable inside quotation marks is sufficient to express an ill-defined function.[3] Thus, consider the expression:

'x'

This normally mentions a symbol used as a variable, viz., the (italicized) 24th letter of the alphabet. But suppose instead that the variable inside the quotes is substitutable, so that it is functionally equivalent to a "blank" that can be filled in by any term. Then, it allows us to define an intensional function i as follows:

$i(x)$ = 'x'

The intensionality comes out when using different terms for 0:

$i(a_0)$ = 'a_0'

$i(a_1)$ = 'a_1'

So to preclude this, a quoted variable is regarded as a name for the variable, nothing more.

Notwithstanding, an analogous ill-defined function can be re-created just by exploiting the possibilities for self-reference. Suppose we define a symbol '*' so that when it precedes a constant, it denotes that very constant. In more detail, '*' could be defined thus:

(D1) $*a_n$ = the constant in the expression $\ulcorner *a_{\underline{n}} \urcorner$

If there are no constraints on self-reference, such an expression would be permitted and is definable more simply as:

[3] This connection between substitution into quotes and an ill-defined function was observed already by, e.g., Tarski (1933/1983: 161).

(D2) $^*\mathrm{a}_n$ = the nth constant

But either way, the *-operation will be ill-defined on 0 in exactly the same way as $i(x)$. Consider that:

(1) $^*\mathrm{a}_0 = $ ‘a_0’

(2) $^*\mathrm{a}_1 = $ ‘a_1’

As with i, the output varies between ‘a_0’ and ‘a_1’, depending on which name for the input used. Concurrently, we can derive absurdities. Suppose that on the lefthand side of (2), ‘a_0’ replaces ‘a_1’, resulting in:

(3) $^*\mathrm{a}_0 = $ ‘a_1’

This along with (1) would mean (by the transitivity of identity):

(4) ‘a_0’ = ‘a_1’

N.B., the substitution needed to derive (3) is *not* a substitution inside quotation marks.

Regardless, since ill-defined functions are excluded, a classical system will not allow definitions like (D2). But contra the usual attitude, this means that some forms of self-reference create problems, even if the language is free of semantic terminology.

3. Discussion

Some additional clarifications are in order. First, it might be asked whether an expression like ‘$^*\mathrm{a}_0$’ is truly self-referential. My inclination is to say ‘yes’, given that it denotes the very constant that is a proper part of the expression. If preferred, however, ‘*’ could be defined to denote not just a proper part but rather the whole expression instead:

(D3) $^*\mathrm{a}_n = \ulcorner {}^*\mathrm{a}_{\underline{n}} \urcorner$

Working with (D3), one could show that '*' is intensional in a similar manner as above.

Second, it might be said that the definitions at hand are inappropriate due to the use of the variable in subscript position. But while a subscripted variable is one way to create a problematic self-reference, it is not the only way (see section 5). Besides, classical logicians otherwise have need of variables in subscript position, e.g., to introduce infinitely many constants as at (ii) above. Variables in subscript position occur elsewhere too, as in some presentations of the anti-diagonal function in Cantorian arguments. (This is not to suggest a problem with such presentations; the point instead is that our argument requires only standard formal devices.)

It may be objected that the preceding shows just that the relevant definitions are simply illegitimate. But that is the point: Such things must be excluded. Yet since ill-defined functions are already *verboten*, we need not impose any novel restrictions. This is why the argument does not have any revisionary implications. At the same time, special attention is called for when commonplace formal devices give rise to an ill-defined function. This is how one might regard the expression 'i' from section 2. Quotation seems unremarkable at first glance; it is just a convenient way to introduce metalinguistic terms. But substitution inside quotation makes for an ill-defined function, and so, a word of warning is required.

Similarly, we have seen that Kripke (like many logicians) seems relatively unconcerned about self-reference. But unrestricted self-reference suffices for an ill-defined function, even in a language free of semantic terms – and so, an explicit advisory seems apt. Put differently, the ban on ill-defined functions implies certain restrictions on self-reference; even though this appears not to be acknowledged.

(Aside: The kinship between the quotation and the *-operation is not coincidental; the latter is a limiting case of the ill-defined operation i. At the same time, it becomes apparent that quotation has a self-referential aspect: A quotation effectively means "the very expression occurring within the single quotes.")

4. Intensionality with Gödel numbers

In conversation, Tim Button has worried that if unrestricted self-reference suffices for absurdity, then Peano Arithmetic might be shown unsound via the method of Gödel numbering. For Gödel numbering enables something functionally like self-reference, and unrestricted self-reference now appears sufficient for intensional inconsistency.

A derivative argument is indeed readily available using Gödel numbers, but it too has no revisionary import. Peano Arithmetic is certainly in no danger. To illustrate, suppose that $g(\epsilon)$ is a function that takes in an expression ϵ and outputs its Gödel number. Then, where $f(n)$ is the nth constant, define the symbol '$^{\#}$' as follows:

(D4) $^{\#}\mathrm{a}_n = g(f(n))$

This defines '$^{\#}$' in such a way that it denotes the Gödel number for the very constant that follows '$^{\#}$'. Thus:

$^{\#}\mathrm{a}_0 = g(\text{‘}\mathrm{a}_0\text{’})$

$^{\#}\mathrm{a}_1 = g(\text{‘}\mathrm{a}_1\text{’})$

But $^{\#}$ determines different Gödel numbers here, even though the input to $^{\#}$ is the same.

Thus, '$^{\#}$' does not express a well-defined operation, and the above definition of '$^{\#}$' is therefore precluded. One might have thought, however, that it is acceptable to define an expression that denotes whichever Gödel number you please. But as our definition shows, '$^{\#}$' cannot be rigged to denote the Gödel number for the very constant occurring after '$^{\#}$'.

However, the argument shows only that Gödel numbering in the metalanguage cannot be used with abandon. It hardly implies that the arithmetical object language is defective. (Nothing in Peano Arithmetic functions like self-reference prior to Gödel numbering.) Granted, since any recursive function is numeralwise expressible in Peano Arithmetic, it may seem that that if contradiction is provable using a Gödel numbering function, then contradiction is provable in the arithmetical object language. Yet the Gödel numbering operation at (D4) *is not a well-defined function*, much less a recursive one. So, formal arithmetic is in no danger.

It remains true that self-reference—and the coding analog of self-reference—cannot be completely unrestrained. But this hardly means that every case of such a thing leads to pathology. (An analogy: The intensional expression 'i' from section 2 does not show all use of quotation to be illegitimate.) The point is just that *if* Gödel codes are abused in a certain way, the proviso against ill-defined functions will be violated.[4]

5. Variations on a theme

In conversation, Gabe Dupre observes that self-reference is not needed for the kind of intensionality at issue. After all, using a variable in subscript position, the following could instead define the intensional expression '*' and no self-reference would seem implicated:

(D5) $*\mathrm{a}_n = n + 57$

In fact, it is tendentious to say that self-reference is absent; after all, the operation adds 57 to the index of the very constant appended to '*'. But no matter: The point remains that the self-referential operations defined at (D1) – (D4) also generate intensional inconsistency. That already shows that self-reference alone can be enough for an intensional context.

What is more, the intensionality can arise without variables in subscript position, as I shall now illustrate. Yet for this, we shall require an approach to definitions that is not always employed, although it is standard in Tarskian semantics.

The relevant approach is one where a functional expression is defined not by using '=' in the object language, as at (D1), but rather by using the denotation-relation in the metalanguage, akin to how an infinitely many number of constants were introduced at (ii). In Tarskian semantics, the function symbols are defined among the metalanguage base clauses;

[4]It would be a gross *non-sequitur*, moreover, to suggest that Gödel's theorems (or any other theorems) depend on abusive coding operations. The standard coding schemes do not use any of the intensional tricks described here.

these are then used in connection with recursive clauses to determine the truth conditions of object-language wff. Thus, the Tarskian base clauses might include the following, where α is any metalinguistic term:

$\ulcorner q(\alpha) \urcorner$ denotes $\ulcorner$‘α’$\urcorner$

Given any expression, this outputs the quotation of the expression. E.g., if ‘a’, ‘b’, ... ‘t’ are constants, then q(‘a’) = ‘ ‘a’ ’, q(‘b’) = ‘ ‘b’ ’, etc. (The function q is well-defined; it is not the intensional operation i from section 2.)

The key observation, then, is that unrestricted self-reference would allow the following to be defined into the object language. Where τ is any singular term (metalinguistic or not):

(D6) $\ulcorner * \tau \urcorner$ denotes τ

Given a singular term τ, the concatenation of ‘*’, followed by τ, is interpreted to denote τ itself. Again, if self-reference were unconstrained, such a definition would be thus far permissible. But suppose that a = b. Then, (D6) implies:

*a = ‘a’

*b = ‘b’

Even though the input is the same, the output differs depending on which term for the input is used. Unrestrained self-reference is again sufficient by itself for an intensional expression, and in this case, it is made possible without a variable in subscript position.

6. Conclusion

To repeat, existing formal systems are free of these intensional cases. Even so, it is worth acknowledging how self-referential operations require *some* restraint in classical logic, especially because self-reference is usually regarded as quite unproblematic.

Acknowledgements

I thank Bradley Armour-Garb, John G. Bennett, Matteo Bianchetti, Ethan E. Brauer, Tim Button, David Chalmers, Morgan Davies, Gabe Dupre, Anil Gupta, Thomas Hofweber, Cory Johnson, Yael Loewenstein, William Lycan, David McCarty, Jay Newhard, Bryan Pickel, Alexander Pruss, Michael Resnik, Ellie Ripley, and Keith Simmons for discussion on various versions of this material. I also thank an audience at the 2018 Central Division meeting of the American Philosophical Association and at the 2022 meeting of the Canadian Philosophical Association, and at the 2022 Joint Session of the Aristotelean Society and the Mind Association.

References

[1] Gödel, Kurt. (1931). 'Über Formal Unentscheidbare Sätze der *Principia Mathematica* und Verwandter Systeme I,' *Monatshefte für Mathematik Physik* 38: 173–198. Reprinted in Gödel, K. (1986). *Collected Works I. Publications 1929–1936*. S. Feferman et al. (eds.), Oxford: Oxford University Press, pp. 144–195.

[2] Kripke, Saul. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72, 690–716.

[3] Priest, Graham. (2008). *An Introduction to Nonclassical Logic: From Ifs to Is* (2nd ed.). Cambridge: Cambridge University Press.

[4] Quine, Willard. (1951). *Mathematical Logic* (Rev. ed.). Cambridge, MA: Harvard University Press.

[5] Quine, Willard. (1960). *Word and Object*. Cambridge, MA: MIT Press.

[6] Tarski, Alfred. (1933/1983). Pojęcie prawdy w językach nauk dedukcyjnych. Towarzystwo Naukowe, Warsaw. Reprinted as: The concept of truth in formalized languages. In J. H. Woodger & J. Corcoran (Eds.), *Logic, Semantics, and Metamathematics: Papers from 1923–1938* (2nd ed., pp. 152–269). Oxford: Oxford University Press.