Logique & Analyse **139-140** (1992), 335-340

SELF-REFERENCE WITH OPERATORS

B.H. SLATER

Patrick Grim, in a recent article,([1]) reflects on Montague's seeming proof that necessity and other modalities cannot be treated as predicates. For self-referential expressions, in that case, would appear to lead inescapably to contradiction. The recent wisdom stemming from this result, Grim realises, has been that necessity and other modalities must be operators instead, but Grim's purpose is to show that there are comparable difficulties with self-reference in certain operator cases, as well.

I propose to show in this paper that Grim does not show there are comparable difficulties with self-reference, in *any* operator case. Moreover, there are in fact no such difficulties, in the cases he is centrally concerned with, for a very simple reason: there is no relevant self-reference in that sort of case, at all. Hence the traditional problems with this notion, in this specific connection, disappear. I comment on one consequence of this general point for our understanding of the natural language operator 'It is provable that'. The general point puts Gödel's incompleteness theorems into a fresh light, since they were concerned, instead, with a certain 'provability' predicate.

The main contrast between operators and the associated predicates I shall mention now: it has not always been observed, but it will be central in what follows. Operators, so to speak, require an interpretation to be built into what they operate on, unlike the associated predicates, which are predicates just of uninterpreted strings of symbols. Thus 'It is provable that 2+2=4' speaks not of *the formula* '2+2=4', but *the fact* that 2+2=4, i.e. what the formula may be used to state about the standard model. This holds simply because in such operator expressions the contained sentence is used, not mentioned. Grim, we shall find, has forgotten this, in what follows.

([1]) 'Operators and the Paradox of the Knower', *Synthese* 94, 1993, 409-28.

1

Now Grim's result, although notionally about such operators as 'It is necessary/It is known that', is significantly limited to another case. Grim is concerned[2] with operators which satisfy, amongst other things

$B(A'') \supset A$,

not

$B(A) \supset A$,

where A" is the 'counterformula' to A on any of various encodings. If A is the i-th element in some enumeration (f) of all formulae in the system, then its counterformula A" is the i-th element in an enumeration (c) of all closed formulae in the system. But, note, there will be no link *in content* between A and A", i.e. any f_i and c_i, in general.

Grim shows that a diagonal lemma (fixed point theorem) is demonstrable in the counterformula case:[3] he shows that, for any formula B(q) of the language, containing just the propositional variable q free, there is a sentence f_k such that

$\vdash f_k \equiv B(c_k)$,

where, of course, c_k is the counterformula to f_k. Using this diagonal lemma he is then able to generate a contradiction.

The first point to note, however, is that this result is not concerned with *self-reference*. For, as before, the contents of the formulae f_k and c_k need not be connected in any way, since all that relates them is their corresponding places in the two orderings of formulae. But Grim says[4] that he is tempted to read B(A") as, say, 'the formula with counterformula A" is necessary', so that, in some informal sense, it says the same thing as B(A). As we saw at the beginning, however, B(A") is not a (predicative) statement about *the formula* 'A"', but an operator expression about its content. Hence B(A") merely says that it is necessary that A" - if B(A) says that it is necessary that A - and the above 'diagonal lemma' does not, in any sense, refer one thing to itself. Moreover, as a supposition about some relation between the contents of A" and A, the axiom

$B(A'') \supset A$

would have to be extraordinarily unlikely, given the arbitrariness of the two

(2) op. cit. p422.

(3) op. cit. p419.

(4) op, cit. pp422,426

orderings of the formulae. So while Grim's mathematical result stands, its significance is severely limited, not just with regard to self-reference, but also in more general respects.

More importantly, since what

$B(A") \supset A$

expresses does not have any relevant sense, for Grim's kind of result to have any bearing on problems to do with self-reference with operators it would have to be replicated in such a case as where

$B(A) \supset A.$

And Grim is satisfied(5) that an operator for which this holds can be consistently included in his system. Indeed, it is obvious that with such an axiom included, in a normal modal system, there *cannot* be troubling self-reference, since

$-B(A \equiv -B(A))$

is a simple theorem of KT. Hence, as was stated at the beginning, the traditional problems with this notion, in this specific connection, simply disappear.

But the question which then arises is why a diagonal lemma in some such self-referential form as

$\vdash f_k \equiv B(f_k)$

is not provable. We must see clearly why it is not.

2

Now in this area, for a start, one clear difference between operators and predicates of the associated formulae is that certain standard means of obtaining self-reference are quite ruled out. For with some naming device for sentences, such as a Gödel numbering, or even just quotation marks, there is no difficulty in including some name of a sentence within itself:

1 = '1 is in English'

is sufficient to illustrate the possibility. Likewise for other referring phrases, which locate the sentence in other ways. But when such a name or referring phrase is part of the same, or a further sentence, then that sentence must naturally be completed *with a predicate* - as 'is in English' above.

The situation with operators is fundamentally different from this, since, if the larger sentence is to contain, at the comparable place, an operator

(5) op. cit. p422,426.

rather than predicate, then it is not *the name of a sentence*, or other similar referring phrase, which completes the larger sentence. It must be a sentence itself, as with the 'p' in e.g.

'q' = 'It is not necessary that p'.

That is so because of the point about use and mention raised at the beginning. But 'q' here *cannot* be 'p', since that would require not the name of the sentence, but *the whole sentence* to be a proper part of itself. Hence the usual way of securing self-reference, in the predicate case, is not available in any operator case.

But 'self-reference', in an extended sense, is not ruled out entirely with operator forms, since, for instance, with 'B' the truth operator 'It is true that', we must have

$\vdash p \equiv B(p)$

quite generally.(6) So why cannot Grim establish a properly self-referential, diagonal lemma, (in this sense) for arbitrary 'B', in parallel to his construction for counterformulas? Grim cannot do this, otherwise his system would be inconsistent upon the addition of the axiom '$B(A) \supset A$'. But *why* cannot Grim do this?

Grim's construction in the counterformula case goes (in effect) as follows:(7) he takes a formula f_j with just one free propositional variable, and defined so that

$\vdash f_j(c_i) \equiv B(f_i(c_i))$".

He then substitutes c_j for c_i, to produce a (closed) formula f_k, equivalent to $f_j(c_j)$, and so equivalent to $B(f_j(c_j))$". This enables him to show

$\vdash f_k \equiv B(c_k)$,

as before. But clearly, if the counterformula relation is replaced by propositional identity, to try to secure something more like proper self-reference, then this construction cannot even get started. For if c_j were identical with f_j it would not be closed, and so could not serve to encode the function $f_j(p)$(8).

(6) The bearing of this point on Tarski's Theorem is given at the end.

(7) op. cit. p419.

(8) Thus it is not an accident that there is in general no semantic link between f_i and c_i: certain semantic links are quite ruled out (and see note 5)

3

Seeing the specific relevance of this matter to Gödel's incompleteness results starts from the realisation that Gödel's notion of provability (as now explicated in the modal system KG4([9])) was meta-mathematical, and hence predicative. But the natural notion of provability is an operator, because of the point about interpretation made at the beginning. In fact Gödel himself realised that the natural notion of proof obeyed the axioms of another modal system, S4, i.e. KT4.([10]) So it is central to the natural notion that

-B(p ≡ -B(p)),

since this is provable even in KT, as we saw before. But this means that the self-reference available to Gödel as a basis for his incompleteness theorems *cannot* be available, even in a transmuted form, in the non-syntactic, contentful, natural language case. where, as before, 'It is provable that 2+2=4', for instance, speaks not of a formula, but a fact.

Now the discussion of Grim's result above has explicated in more detail just why the kind of self-reference provably unavailable for modalities satisfying KT is even not grammatically generable. And thus it shows a Wittgensteinian point, about how *irrelevant* to the understanding of our natural, non-formalist, notion of proof are Gödel's theorems, in a central sense. In fact it can be proved in K4, and so in both KG4 and KT4, that the following theorems hold:([11])

B(p ≡ -B(p)) ⊃ (B(p) ⊃ B(q)),

B(p ≡ -B(p)) ⊃ (B(-B(q)) ⊃ B(r)),

So, if we can detach the antecedents, for some specific p, we have versions of Gödel's two theorems: if that particular p is provable then anything is (i.e. the system is inconsistent), and if consistency is provable (proving the system is consistent means proving there is something it cannot prove), then it is, in fact, inconsistent.

Evidently 'proving' something does not invariably require that it be true - only if axiom T is involved does that consequence automatically follow - but without T, i.e. for modalities in KG4 (KTG4 is inconsistent) the oportunity is opened up to establish, and therefore detach, the antecedents above.

([9]) Boolos,G. 1979, *The Unprovability of Consistency*, C.U.P.

([10]) Gödel,K. 1969, 'An Interpretation of the Intuitionistic Sentential Logic', in *The Philosophy of Mathematics*, ed, J.Hintikka, O.U.P.

([11]) Smullyan,R. 1987, *Forever Undecided*, O.U.P., p102f.

But clearly we cannot detach the antecedents if we are dealing with the natural notion of provabilty, because of the impossibility of relevant self-reference, in that case, as before. Hence Gödel's theorems do not follow in the operator case.

A similar point may be made, even more readily, with respect to Tarski's theorem. For a truth operator is straightforwardly definable within normal modal systems, as mentioned above. Tarski, as is well known, took truth to be predicative, and it is in that context that his theorem arises. The discrepancy shows his approach must fail to capture the natural notion of truth, in line with Montague's result.[12]

The University of Western Australia.

(12) See also Prior,A.N. 1971, *Objects of Thought*, O.U.P., Ch. 7, and, for instance, Slater,B.H. 1986, 'Prior's Analytic', *Analysis*, and Slater,B.H. 1991, 'Liar Syllogisms and Related Paradoxes', *Analysis*.