PATRICK GRIM

OPERATORS IN THE PARADOX OF THE KNOWER[1]

ABSTRACT. Predicates are term-to-sentence devices, and operators are sentence-to-sentence devices. What Kaplan and Montague's Paradox of the Knower demonstrates is that necessity and other modalities cannot be treated as predicates, consistent with arithmetic; they must be treated as operators instead. Such is the current wisdom.

A number of previous pieces have challenged such a view by showing that a predicative treatment of modalities need *not* raise the Paradox of the Knower. This paper attempts to challenge the current wisdom in another way as well: to show that mere appeal to modal operators in the sense of sentence-to-sentence devices is insufficient to *escape* the Paradox of the Knower. A family of systems is outlined in which closed formulae can encode other formulae and in which the diagonal lemma and Paradox of the Knower are thereby demonstrable for operators in this sense.

Predicates are term-to-sentence devices: functions that take the terms of a language as input and render sentences (or open formulae) as output. Sentential operators or connectives, on the other hand, are sentence-to-sentence devices: they take sentences (or open formulae) as input and render sentences (or open formulae) as output. Such at least is the current wisdom. Quine's 'Nec' is intended as a predicate. The familiar '$\Box$' of modal logic, in contrast, is an operator.[2]

Another claim that appears as part of the current wisdom is this: that what Kaplan and Montague's Paradox of the Knower demonstrates is that necessity and other modalities cannot be treated as predicates, consistent with arithmetic. They must be treated as sentential operators, instead.[3] In Montague's words,

> if necessity is to be treated syntactically, that is, as a predicate of sentences . . . then virtually all of modal logic, even the weak system S1, must be sacrificed.
>
> This is not to say that the Lewis systems have no natural interpretation. Indeed, if necessity is regarded as a sentential operator, then perfectly natural model-theoretic interpretations may be found. (Montague [1963]/1974, p. 294)

In Putnam's words,

> there is no paradox associated with the notion of necessity as long as we take the '$\Box$' as a statement connective (in the degenerative sense of 'unary connective') and not – in spite of Quine's urging – as a predicate of sentences . . . (Putnam [1967]/1983, p. 308)[4]

Synthese **94**: 409–428, 1993.

A number of results indicate that this current wisdom regarding predicates, operators, and the Paradox of the Knower is at least seriously incomplete. Brian Skyrms (1978), Tyler Burge (1978, 1984), C. Anthony Anderson (1983), and J. des Riviéres and H. Levesque (1986) have pointed out that not all treatments of necessity and similar modalities as predicates will fall victim to the Knower. Limit the expressive power of such predicates in certain ways, in particular, and they will mimic the behavior of standard operators enough to escape the Paradox of the Knower.[5]

In what follows I want to sketch some results which suggest that the current wisdom is incomplete in another respect as well. Just as predicative treatments of modality don't necessarily fall victim to the Paradox of the Knower, sentential operators don't necessarily escape it. In systems with certain expanded expressive resources, in particular, the diagonal lemma and the Paradox of the Knower will not be limited to predicates: they will be demonstrable for sentential operators as well.

In Sections 1–3, I introduce a family of systems around a conveniently simple example Q+. Relative consistency is demonstrated for Q+ in Section 4. In Section 5, I offer a form of the diagonal lemma for such systems, and in Section 6 construct the Paradox of the Knower in terms of operators. Some generalizations are offered in conclusion.

Here a reservation regarding terminology should be noted, however. As indicated above, the work of Skyrms through des Riviéres and Levesque relies on the construction of systems in which certain predicates behave enough like standard operators to escape the Paradox of the Knower. The core of the work that follows is the construction of systems in which certain sentential operators have enough of the character of predicates to fall victim to the Paradox of the Knower. But it might then be charged that des Riviéres and Levesque's 'predicates' are merely operators in disguise, or that my 'operators' are really predicates.

There is, I think, something to be said for this way of characterizing both previous work and that which follows. It should be noted, however, that such an approach would still demand that we sacrifice a part of the common wisdom. If des Riviéres and Levesque's predicates are to be written off as operators or my operators are to be written off as predicates, it's clear that the distinction between predicates and operators can't be as clear or clean as the common wisdom seems to suppose.

In particular, predicates and operators can't be said simply to be term-to-sentence and sentence-to-sentence devices, respectively.

More fundamentally, what the work that follows seems to indicate is that the essential difference between systems which fall victim to the Paradox of the Knower and those which avoid it is not simply the syntactical form of functions such as 'Nec' and '$\Box$'. The essential difference is, rather, the expressive power that systems at issue allow such functions, whatever their syntactical form.

1. SYSTEMS AT ISSUE: INITIAL SPECIFICATIONS

In what follows I want to consider standard systems broadened to include (1) a restricted form of propositional quantification and identity, and (2) additional axiomatic machinery for certain expanded powers of representation. The particular restrictions on propositional quantification and identity at issue and the purpose and form of the additional axiomatic machinery will be specified below.

Here it's convenient to start with a particularly simple example, Q+, built on the foundation of system Q of Robinson arithmetic. It should be noted, however, that a broad range of systems are at issue and that basic results do not ultimately depend on the use of systems of number theory in particular. With minor variations noted at certain points, all basic results will hold, for example, for systems dealing entirely with sentential matters.

In order to construct the sample system Q+ we start with the standard symbolism and grammar for Q. To the logical symbols of Q we add propositional variables $p_1, p_2, p_3, \ldots$. To the non-logical symbols we add a two-place propositional operator $\bigcirc$, which will eventually be used to express diagonalization. To the standard grammar for Q we add the following:

Let a *propositional expression* be any closed formula or any propositional variable, and only these.

(G1) If e_1 and e_2 are propositional expressions, $(e_1 = e_2)$ is a wff.

(G2) If e_1 and e_2 are propositional expressions, $\bigcirc(e_1, e_2)$ is a wff.

(G3) If P is a wff and p is a propositional variable, then $\forall p(P)$ is a wff.

(G4) There are no other wffs.[6]

I envisage the standard axioms and rule for Q in the following form:

For any wffs P, Q, and R, any individual variables x and y, and any term t:

I. Propositional schemata
(A1) $(P \rightarrow (Q \rightarrow P))$.
(A2) $((P \rightarrow (Q \rightarrow R)) \rightarrow ((P \rightarrow Q) \rightarrow (P \rightarrow R)))$.
(A3) $((\sim Q \rightarrow \sim P) \rightarrow (P \rightarrow Q))$.

II. Schemata for first-order quantification
(A4) $(\forall x(P) \rightarrow P(t/x))$, where $P(t/x)$ is the result of substituting in P term t for all free occurrences of x, and where t is a term free for x in P.
(A5) $(P \rightarrow \forall x(P))$, where x does not occur free in P.
(A6) $(\forall x(P \rightarrow Q) \rightarrow (\forall x(P) \rightarrow \forall x(Q)))$.
(A7) $\forall x(P)$, where P is an axiom from (A1)–(A7), inclusive.

III. Schemata for first-order identity
(A8) $\forall x(x = x)$.
(A9) All closures of: $((x = y) \rightarrow (P \rightarrow Q))$, where Q is just like P except that at one or more places where x occurs free in P, y occurs free in Q.

IV. Schemata for arithmetic
(A10) $\forall x \forall y((x' = y') \rightarrow (x = y))$.
(A11) $\forall x \sim(0 = x')$.
(A12) $\forall x(\sim(x = 0) \rightarrow \exists y(x = y'))$.
(A13) $\forall x(x + 0 = x)$.
(A14) $\forall x \forall y(x + y') = (x + y)')$.
(A15) $\forall x(x \cdot 0 = 0)$.
(A16) $\forall x \forall y(x \cdot y' = (x \cdot y) + x)$.

RMP (modus ponens): Q may be inferred from $(P \rightarrow Q)$ and P.

The axioms and rule for Q+ can then be specified as follows. We take all of the above as applicable to wffs of Q+, except that (A7) must be modified to read:

(A7′) $\forall x(P)$, where P is an axiom from (A1)–(A7′) or (A17)–(A20), inclusive.[7]

We now add:

For any wffs P, Q, and R, any individual variables x and y, any propositional variables p and q, any term t, and any propositional expression e:

V. Schemata for propositional quantification

(A17) $(\forall p(P) \rightarrow P(e/p))$, where $P(e/p)$ is the result of substituting in P a propositional expression e for all free occurrences of p, where e is free for p in P.[8]

(A18) $(P \rightarrow \forall p(P))$, where p does not occur free in P.

(A19) $(\forall p(P \rightarrow Q) \rightarrow (\forall p(P) \rightarrow \forall p(Q)))$.

(A20) $\forall p(P)$, where P is an axiom from (A1)–(A7′) or (A17)–(A20), inclusive.

VI. Schemata for propositional identity

(A21) $\forall p(p = p)$.

(A22) All closures of: $((p = q) \rightarrow (P \rightarrow Q))$,[9] where Q is just like P except that at one or more places where p occurs free in P, q occurs free in Q.

Here the character of restrictions on propositional quantification and identity should be clear: by (A17), only propositional variables and closed formulae are permitted in instantiation for propositional variables. Thus Q+ is to this point an extension of Q with propositional quantification and identity, instantiation for propositional variables being restricted to closed formulae and propositional variables. In general we will consider systems broadened to include propositional quantification and identity restricted in this sense.

In order to complete Q+, we must also include axiomatic machinery for certain expanded powers of representation. Let me first sketch a notion of functions of formulae and a general strategy for representing effectively calculable functions of formulae within a system such as Q+. I will then introduce the particular function of formulae that is to be represented within Q+.

2. REPRESENTING FUNCTIONS OF FORMULAE

Recursive functions of numbers, representable in Q, are of course familiar.[10] Here we introduce a similar notion of functions of formulae – in particular, functions of closed formulae (henceforth also 'c-formulae'), which take closed formulae as arguments as values. We will be concerned in particular with effectively calculable functions of c-formulae.

We will say that an n-place function f of c-formulae is *represented* in a system T if there is a formula $D(p_1, \ldots, p_n, p_{n+1})$ such that for any closed formulae $c_1, \ldots, c_n$, c, if $f(c_1, \ldots, c_n) = c$, then

$$\vdash_T \forall p_{n+1}(D(c_1, \ldots, c_n, p_{n+1}) \equiv p_{n+1} = c).^{11}$$

In what follows only one-place functions of c-formulae will in fact be used.

We can now introduce a general strategy for building axiomatic machinery in order to represent such functions within certain systems.

We assume a system with at least the syntax, axioms, and rule so far outlined for Q+, and suppose an effectively calculable one-place function f of c-formulae we wish to represent. We also assume some particular Gödel numbering for all formulae of the system. To each closed formula c_1, as argument for f, and each c_2, as value, will thus correspond Gödel numbers $\overline{c_1}$ and $\overline{c_2}$.

Consider now a procedure that starts with a number $\overline{c_1}$, decodes it as a formula c_1, calculates our function f for c_1 as argument, giving us a formula c_2, and encodes c_2 as a number $\overline{c_2}$. Since Gödel encoding and decoding are effective and f has been assumed effectively calculable, this whole procedure will be effective as well.

Given Church's thesis, then, there will be a recursive function f' of numbers such that $f'(\overline{c_1}) = \overline{c_2}$ will hold just in case $f(c_1) = c_2$ for our chosen f.[12] All recursive functions of numbers are represented in Q and are represented by the same formulae in any extension of Q.[13] There will then be a formula D of Q and so of the system at issue such that for any $\overline{c_1}$ and $\overline{c_2}$, if $f'(\overline{c_1}) = \overline{c_2}$, then

$$\vdash \forall x(D(\overline{\mathbf{c_1}}, x) \equiv x = \overline{\mathbf{c_2}}),$$

where $\overline{\mathbf{c_1}}$ and $\overline{\mathbf{c_2}}$ are numerals within the system for $\overline{c_1}$ and $\overline{c_2}$, respectively.

In order to represent our chosen function of formulae f within such a system, let us add as axioms all instances of the following schema:

$$\forall p(\bigcirc(c_1, p) \equiv p = c_2) \equiv \forall x(D(\overline{\mathbf{c_1}}, x) \equiv x = \overline{\mathbf{c_2}}).$$

The left halves of instances of this schema may be thought of as propositional quantifications over closed formulae; the right halves as individual quantifications over their corresponding Gödel numbers. Intuitively the strategy as a whole might then be thought of as 'reading off' (or 'reading up') representation of an effectively calculable function of c-formulae from representation of a corresponding numerical function.

For any c-formulae c_1 and c_2, if $f(c_1) = c_2$, the following will now be a theorem of our system:

$$\forall p(\bigcirc(c_1, p) \equiv p = c_2),$$

and thus our chosen function f will be represented. Although we have confined ourselves here to a one-place function f, the same strategy will clearly serve for any n-place function of c-formulae.

On the basis so far outlined, we might choose to introduce axiomatic machinery for representation of any effectively calculable function of closed formulae. In building the particular system Q+, I want to introduce just one such function.

Although we have here used the familiar number-theoretic capacities of our basic system Q in order to introduce representation of functions of closed formulae, it should be noted that it is the latter that is important for the results that follow. Despite our use of Q+ as a convenient example throughout, those results are thus not tied directly to number theory; they will hold as well for systems in which effectively calculable functions of closed formulae are represented directly rather than by way of representation of numerical functions. For these, of course, a simpler axiomatization directly in terms of $\bigcirc$ would be possible.

3. COUNTER-FORMULAE AND DIAGONALIZATION

We continue to consider as an example a system with the syntax and axioms so far outlined for Q+. By Gödel numbering, of course, all formulae of such a system can be recoverably encoded as numerals. For a broad range of systems, however, it is also the case that all formulae can be recoverably encoded as closed formulae.

In general we assign the *rth* element c_r of a chosen enumeration of c-formulae to the *rth* element f_r of a chosen enumeration of all formulae as the latter's *counter-formula*.[14] Given any pair of such enumerations, each formula will be recoverably encoded as a closed formula much as it is more standardly encoded as a Gödel number within standard systems.[15]

For our sample system Q+, we will assume as given a particular c-formula encoding of this type as well as a particular Gödel numbering.

For the moment we will indicate the counter-formula of a formula f_j as c_j. At a later stage an alternative symbolism proves more perspicuous; there we will indicate a counter-formula by double overlining, such that $\bar{\bar{A}}$ will stand in for the counter-formula of A.

For a formula f_j and its counter-formula c_j, let us specify the c-diagonalization of f_j as the expression:

$$\exists p(p = c_j \,\&\, f_j).$$

Informally, if f_j contains just the propositional variable p free, the c-diagonalization of f_j will be a sentence that 'says that' f_j holds of its own counter-formula c_j.

Consider now a function $diag^c$ of closed formulae such that if the c-diagonalization of a formula f_j is a formula f_k, $diag^c(c_j) = c_k$. We will refer to $diag^c$ for fairly obvious reasons as a c-diagonalization function on c-formulae.

Clearly $diag^c$ is effectively calculable, as are Gödel encoding and decoding. Following the reasoning of the preceding section, then, there will be an effectively calculable numerical function d such that for any c-formulae c_1 and c_2 and corresponding Gödel numbers $\overline{c_1}$ and $\overline{c_2}$, $d(\overline{c_1}) = \overline{c_2}$ will hold just in case $diag^c(c_1) = c_2$.[16] By Church's thesis, d will be recursive, and so will be represented in Q and by the same formula in any extension of Q.

For some formula D of Q and of the system at issue, then, for any $\overline{c_1}$ and $\overline{c_2}$ such that $d(\overline{c_1}) = \overline{c_2}$, the following will be a theorem:

$$\forall x(D(\overline{\mathbf{c_1}}, x) \equiv x = \overline{\mathbf{c_2}}).$$

In order to represent $diag^c$, given syntax and axioms so far specified for Q+, we add all instances of the following axiom schema:

VII. Schema for representing $diag^c$

(A23) $\forall p(\bigcirc(c_1, p) \equiv p = c_2) \equiv \forall x(D(\overline{\mathbf{c_1}}, x) \equiv x = \overline{\mathbf{c_2}})$.

This final axiom schema completes the specification for Q+, which can now be characterized as an extension of Q: (1) with propositional quantification and identity, instantiation for propositional variables restricted to closed formulae and propositional variables; and (2) in which a c-diagonalization function on c-formulae is represented.

Though specified so as to incorporate all machinery necessary for the proofs that follow, Q+ is merely intended as an example from a broader family of systems. As noted above, the number-theoretical capacities we have used to introduce certain features of Q+ are not strictly necessary; (A23) might be replaced with a simpler schema for diagonalization dealing directly in terms of c-formulae and $\bigcirc$, for example. In general, we will be concerned with a range of systems for which the two conditions above hold.

4. CONSISTENCY

One of the advantages of Q+'s simplicity is that it affords us a fairly simple relative consistency proof: if Q is consistent, so is Q+.

Here the basic strategy is to introduce a 'dump' translation '$\downarrow$' which takes wffs of Q+ into wffs of Q. '$\downarrow$' converts wffs of Q+ to those of Q essentially by dumping formulae with propositional quantification, identity, and $\bigcirc$ down to formulae with simple individual quantification, identity, and formula D of Q. For the sake of brevity I omit details.

Using such a translation it can be shown that:

(1) For any $P \in \mathrm{WFF}_{Q+}$, if P is an axiom of Q+, then $\vdash_Q \downarrow (P)$;

(2) For each P, Q, $\in \mathrm{WFF}_{Q+}$, if $\vdash_Q \downarrow (P \rightarrow Q)$ and $\vdash_Q \downarrow (P)$, then $\vdash_Q \downarrow (Q)$,

and thus

(3) For any $P \in \mathrm{WFF}_{Q+}$, if $\vdash_{Q+} P$, then $\vdash_Q \downarrow (P)$.

Contradictory wffs of Q+ become contradictory wffs of Q under '$\downarrow$'. Were Q+ inconsistent, therefore, Q would be as well.[17]

Consider now a system $Q + \triangle$, which is essentially Q+ with syntax broadened to include a one-place operator '$\triangle$' on closed formulae and propositional variables. We modify the grammar specified for Q+ above by deleting our previous (G4) and adding:

(G4) If e_1 is a propositional expression, $\triangle(e_1)$ is a wff.

(G5) There are no other wffs.

(A1)–(A23) can then be rewritten for wffs of $Q + \triangle$.

Let $diag^c$ here be a c-diagonalization function of c-formulae for *this* system. Given a particular Gödel numbering, let d be a numerical function such that $d(\overline{c_1}) = \overline{c_2}$ just in case $diag^c(c_1) = c_2$. Let D′ be a formula that represents this d in Q and its extensions, and replace the 'D' of (A23) with this 'D′'. This completes our specification for $Q + \triangle$.

$Q + \triangle$, like Q+, will be provably consistent relative to Q. One form of the proof, in two steps, is as follows.

We introduce a slight variation on Q+, system Q+′, in which all is as in Q+ except that formula D′ of Q rather than formula D appears in (A23). Q+′ will be provably consistent relative to Q using a variant on '$\downarrow$' which uses D′ in place of D.

We also introduce a further translation '$\Downarrow$' that takes wffs of $Q + \triangle$ into wffs of Q+′ by everywhere replacing '$\triangle(e_1)$' for propositional expressions e_1 with '$(e_1 = e_1)$'.[18] It can now be shown that:

(1) For any $P \in WFF_{Q+\triangle}$, if P is an axiom of $Q + \triangle$, $\Downarrow(P)$ is an axiom of Q+′;

(2) For each $P, Q, \in WFF_{Q+\triangle}$, if $\vdash_{Q+'} \Downarrow(P \to Q)$ and $\vdash_{Q+'} \Downarrow(P)$, then $\vdash_{Q+'} \Downarrow(Q)$;

and thus:

(3) For any $P \in WFF_{Q+\triangle}$, if $\vdash_{Q+\triangle}(P)$ then $\vdash_{Q+'} \Downarrow(P)$.

Contradictory wffs of $Q + \triangle$ become contradictory wffs of Q+′ under '$\Downarrow$', and thus if Q+′ is consistent so is $Q + \triangle$.

If Q is consistent, then so is $Q + \triangle$.

5. THE DIAGONAL LEMMA

The diagonal lemma is of course familiar in the following form for systems S in which a numerical diagonalization is representable:[19]

For any formula B(y) of the language of S containing just the variable y free, there is a sentence G such that

$$\vdash_S G \equiv B(\bar{G}),$$

where $\bar{G}$ is the Gödel number of G.

For systems at issue here, the diagonal lemma is also demonstrable in a 'higher' form. For systems with propositional quantification and identity, instantiation for propositional variables restricted to closed formulae and propositional variables:

Let T be a theory in which a c-diagonalization function of c-formulae is represented. Then for any formula B(q) of the language of T, containing just the propositional variable q free, there is a sentence f_k such that

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

where c_k is the counter-formula of f_k.

The proof is as follows. Let '*diag*^{c}' name our c-diagonalization function of c-formulae in T, and let $\bigcirc(p, q)$ represent that function. Then for any closed formulae c_j and c_k, if $diag^c(c_j) = c_k$,

$$\vdash_T \forall q(\bigcirc(c_j, q) \equiv q = c_k).$$

Let formula f_j be: $\exists q(\bigcirc(p, q)\ \&\ B(q))$. f_j is a formula of T with just the propositional variable p free. Its counter-formula will be c_j.

Let f_k be: $\exists p(p = c_j\ \&\ \exists q(\bigcirc(p, q)\ \&\ B(q))$. f_k's counter-formula will be c_k.

Since f_k is logically equivalent to $\exists q(\bigcirc(c_j, q)\ \&\ B(q))$, we will have

$$\vdash_T f_k \equiv \exists q(\bigcirc(c_j, q)\ \&\ B(q)).$$

Since f_k is the c-diagonalization of f_j,

$$diag^c(c_j) = c_k.$$

So: $\vdash_T \forall q(\bigcirc(c_j, q) \equiv q = c_k)$, by the representability of *diag*^{c}.
So: $\vdash_T f_k \equiv \exists q(q = c_k\ \&\ B(q))$.
So: $\vdash_T f_k \equiv B(c_k)$.[20]

Here we should also perhaps note a form of Tarski's theorem. We will term $\bigstar$ a counter-formula truth operator for T if for every sentence f_i of the language of T,

$$\vdash_T f_i \equiv \bigstar c_i.$$

If a system T meeting the specifications above is consistent, it can contain no counter-formula truth operator. For using '$\sim\bigstar$' as 'B' in the diagonal lemma, for some sentence f_a:

$$\vdash_T f_a \equiv \sim\bigstar c_a.$$

But then: $\vdash_T f_a \equiv \bigstar c_a$.
And so: $\vdash_T \bigstar c_a \equiv \sim\bigstar c_a$.

6. THE PARADOX OF THE KNOWER

Consider a system: (1) with propositional quantification and identity, instantiation for propositional variables restricted to closed formulae and propositional variables; (2) in which a c-diagonalization function of c-formula is represented; and (3) the syntax of which includes a one-place operator '$\triangle$' on closed formulae. Q + $\triangle$, for which a relative consistency proof was sketched in Section 4, can serve as a simple example.

Here the argument is more perspicuous if we shift symbolism slightly. Instead of using 'c_j' to indicate the counter-formula of a formula f_j, as above, we will use double overlining to indicate counter-formulae; $\bar{\bar{A}}$ will stand in for the counter-formula of a formula A.

Within a system of the sort specified, '$\triangle(\bar{\bar{A}})$' is intended, perhaps, as 'the formula with counter-formula $\bar{\bar{A}}$ is necessary', '. . . is in principle knowable', '. . . is known by God', or the like. On any of these readings it is tempting to add the following schemata:

(1) $\triangle(\bar{\bar{A}}) \rightarrow A;$
(2) $\triangle(\bar{\bar{T}});$
(3) $I(\bar{\bar{A}}, \bar{\bar{B}}) \rightarrow \cdot\, \triangle(\bar{\bar{A}}) \rightarrow \triangle(\bar{\bar{B}}),$

where T is some convenient tautology and 'I(x, y)' represents the deducibility relation for the system at issue. (For systems other than systems of number theory, (3) can be replaced with a schema giving us $\triangle(\bar{\bar{A}}) \rightarrow \triangle(\bar{\bar{B}})$ as an axiom whenever B is deducible from A.)[21]

With such apparently innocuous additions, however – and despite the fact that '$\triangle$' is here a sentential operator in the sense of a sentence-to-sentence device – our system becomes inconsistent.

The proof is as follows. From the diagonal lemma, with '$\sim \triangle(q)$' for 'B(q)', we have for some sentence S:

$$\vdash S \equiv \sim \triangle(\bar{\bar{S}}) \quad (*).$$

Now:

$\vdash \triangle(\bar{\bar{S}}) \rightarrow S$ by (1)
$\vdash \triangle(\bar{\bar{S}}) \rightarrow \sim \triangle(\bar{\bar{S}})$ by (*)
$\vdash \sim \triangle(\bar{\bar{S}})$
$\vdash S$ by (*)
$\vdash S \rightarrow (T \rightarrow S)$ for any T

$\vdash T \rightarrow S$
$T \vdash S$
$\vdash I(\bar{T}, \bar{S})$
$\vdash I(\bar{T}, \bar{S}) \rightarrow . \triangle(\bar{\bar{T}}) \rightarrow \triangle(\bar{\bar{S}})$ by (3)
$\vdash \triangle(\bar{\bar{\bar{T}}}) \rightarrow \triangle(\bar{\bar{\bar{S}}})$
$\vdash \triangle(\bar{\bar{\bar{T}}})$ by (2)
$\vdash \triangle(\bar{\bar{\bar{S}}})$.

The fourth and last lines give us a contradiction.

7. SOME CONCLUDING OBSERVATIONS

The diagonal lemma and the Paradox of the Knower, then, are not merely a matter for predicates – they will hold for certain sentential operators within certain systems as well.

How far do these results extend, and what do they ultimately have to tell us? Here let me offer a few brief observations

Our basic results above have been outlined using as examples Q+ and Q + $\triangle$: forms of system Q of Robinson arithmetic extended to include: (1) propositional variables in the context of operators $\bigcirc$ and $\triangle$ as well as $=$; (2) in which axioms for propositional identity and propositional quantification and instantiation for propositional variables are restricted to closed formulae and propositional variables; and (3) in which a c-diagonalization function of c-formulae is represented.

It is clear from the basic structure of the argument above, however, that these results can be generalized. As indicated throughout, the number-theoretic capabilities of our basic systems are not strictly necessary; basic results at issue will hold for systems in which a diagonalization function of closed formulae is represented directly, for example.

In Q+ and Q + $\triangle$ it is the closed formulae of the system as a whole that serve as encoding counter-formulae. There is, however, nothing sacrosanct about using all of these; any of various less inclusive classes of closed formulae would do as well. In initially exploring some of the territory above, for example, systems closely related to Q+ and Q + $\triangle$ were used which took as counter-formulae closed formulae exclusively of Q.

In general, any denumerably infinite subset of the closed formulae of such a system can serve the function of encoding. Propositional quantification must of course extend to that class of closed formulae, in the manner of Q+ and Q + $\triangle$, and a counter-formula diagonalization function in terms of that encoding must be represented. Given these, however, we will have essentially all the basic elements required to give us an operator form of the diagonal lemma and the Paradox of the Knower.[22]

It is also important to note some major limitations of these results, however.

In terms of the Paradox of the Knower, what we have shown with regard to systems such as Q+ is that inclusion of an operator $\triangle$ with the following schemata – where $\bar{\bar{A}}$ is the counter-formula for A on any of various encodings – can be expected to lead to contradiction:

$$\triangle(\bar{\bar{A}}) \rightarrow A$$
$$\triangle(\bar{\bar{T}})$$
$$I(\bar{A}, \bar{B}) \rightarrow . \triangle(\bar{\bar{A}}) \rightarrow \triangle(\bar{\bar{B}}).$$

We have not shown, however, that inclusion of a more familiar $\triangle$ operator with these schemata will lead to contradiction in such a system:

$$\triangle(A) \rightarrow A$$
$$\triangle(T)$$
$$I(\bar{A}, \bar{B}) \rightarrow . \triangle(A) \rightarrow \triangle(B).$$

Indeed the latter need not give us a contradiction. Where A and B are restricted to closed formulae, at least – as the grammar of Q + $\triangle$ would demand – an operator with these schemata can be consistently incorporated in an extension of Q + $\triangle$ just as it can be consistently incorporated in simpler extensions of Q.[23]

An operator $\triangle$(A) which functions directly on a non-encoded formula A can thus be consistently included where an operator $\triangle(\bar{\bar{A}})$ on a counter-formula encoding A cannot. This holds, moreover, despite the fact that $\triangle$(A) and $\triangle(\bar{\bar{A}})$ are both sentential operators in the sense of taking sentences as both argument and value, input and output, and despite the fact that it is tempting to read the two operators and, thus, the two sets of schemata above as in some informal sense 'saying' the same thing.[24]

What generalizations can we draw regarding predicates and operators in the diagonal lemma and the Paradox of the Knower?

If in line with current wisdom we take predicates to be simply term-to-sentence devices and take operators to be simply sentence-to-sentence devices, at least, the diagonal lemma and the Paradox of the Knower will not be restricted to predicates: they will appear in full strength for certain operators in certain systems as well.

Whether a sentence-generating function takes terms or sentences as arguments, then – whether it functions as a predicate or sentential operator in this sense – is not what ultimately determines whether it will give us a form of the diagonal lemma and the Paradox of the Knower. What *does* ultimately make such a difference seems to be rather the expressive capacity afforded a function of either type within a given system.

In standard systems of arithmetic, for example, predicative sentences can 'reach around' so as to self-apply in a certain sense (or so as to apply to more complex formulae). Less metaphorically put: among the terms to which a predicate P can apply in such a system is a term x which encodes the sentence Px itself (or which encodes some more complex formula in which Px is embedded). This is of course the core of the diagonal lemma and, hence, the Paradox of the Knower in its standard form: the fact that for a given predicate P there will be a sentence F with a term-encoding $\bar{F}$ such that $\vdash F \equiv P(\bar{F})$. By means of Gödel number term-encoding and within the bounds of material equivalence predicative sentences are given an expressive range that extends in effect to self-application.

Standard operator sentences in standard systems, in contrast, are expressively impoverished in precisely this regard: they cannot 'reach around' to self-apply or to apply to more complex formulae. The range of sentences to which a non-trivial operator $\triangle$ can apply on a given occasion does not extend as far as a sentence A expressing that application $\triangle(A)$ itself. The expressive range of standard operators in familiar systems is thus simply smaller than that afforded predicates in standard systems by means of encoding and representation of certain functions.

As noted in the introduction, J. des Riviéres and H. Levesque (1986) manage to insulate certain predicates from the Paradox of the Knower by explicitly restricting their expressive applicational range to that of familiar operators. Here, in contrast, we have introduced the diagonal lemma and the Paradox of the Knower for certain operators by effectively extending their expressive applicational range: by considering systems in which formulae can serve an encoding function and in which certain functions of encoding formulae are represented.

The general lesson to be drawn seems to be that it is the effective expressive power of a system – whether that power appears in the form of predicates or operators – that lies at the real heart of the diagonal lemma and related results.

NOTES

[1] I am deeply indebted to Robert F. Barnes and Evan W. Conyers, without whom these ideas might not have germinated and certainly would not have grown. Many of the results offered here evolved in the course of mutual discussion and correspondence. I am also grateful to an anonymous reviewer for *Synthese* for many very helpful suggestions.

The current paper contains the technical results promised in Footnote 25 of Grim (1988) and Footnote 26, Chapter 3, of Grim (1991).

[2] As a representative statement of the current wisdom on predicates and operators, see for example Reinhardt (1980). Classic characterizations of the basic distinction between predicates on the one hand and sentential operators or connectives on the other appear in Church (1956, p. 36), and Anderson and Belnap (1975, p. 477). I will use the term 'sentential operator' – or simply 'operator' – in preference to 'sentential connective' throughout.

[3] In a nutshell, the standard Paradox of the Knower is as follows. Start with any system adequate for arithmetic, such as system Q of Robinson arithmetic, and extend the syntax to include a predicate 'K'. For a sentence 'A', '$K(\bar{A})$' is intended perhaps as 'the sentence with Gödel number $\bar{A}$ is necessary', '. . . is in principle knowable', '. . . is known by God', or the like. Richmond Thomason lists other plausible candidates for a predicate 'K': '. . . is certain', '. . . can be demonstrated', '. . . follows from what I know', 'logic alone suffices to establish . . .', and '. . . is trivial' (Thomason 1977, p. 350). With any of these readings it is tempting to add the following schemata to our system, where '$I(\bar{x}, \bar{y})$' represents the deducibility relation for the system at issue:

$$K(\bar{A}) \rightarrow A$$
$$K(\overline{K(\bar{A}) \rightarrow A})$$
$$I(\bar{A}, \bar{B}) \rightarrow . \, K(\bar{A}) \rightarrow K(\bar{B}).$$

With that addition, however, our system becomes provably inconsistent.

The original result appears in Kaplan and Montague ([1960]/1974). A particularly nice exposition appears in Anderson (1983). As indicated in Sections 6 and 7, it is possible to strengthen the Knower slightly: provable inconsistency results even if the second schema is replaced merely with $K(\bar{T})$ for any simple tautology T.

[4] A clear outline of the current wisdom regarding predicates and operators in the Paradox of the Knower also appears in introductory sections of Asher and Kamp (1989).

[5] Skyrms, Burge, and Anderson all limit the expressive power of a predicate (or series of predicates) by imposing hierarchical constraints. Des Riviéres and Levesque introduce predicates the range of which is quite explicitly mapped off corresponding operators.

[6] This may seem a peculiar way to introduce a grammar for propositional quantification; it is not, for example, that employed in Bull (1969), Fine (1970, 1977), and Kaplan (1970), and the wffs of Q+ may seem an amputated set in comparison with those systems. Here, however, my aim is to keep Q+ as simple as possible, and I bring in essentially

only what is required for later proofs. This grammar allows us the fairly simple relative consistency proof for Q+ sketched in Section 4.

'∃' and '&' will be used as standard abbreviations and conventions for parentheses will occasionally be relaxed.

[7] See also note 9, however.

In a more standard form, of course, we might group axioms (A7′) and (A20) at the end.

[8] e is free for p in P iff no free occurrences of p lie within the scope of any quantifier ∀q, where q is a variable in e. Since a variable can occur free in a propositional expression e only if e is a propositional variable, however, we might say simply 'e is free for p in P iff no free occurrences of p lie within the scope of any quantifier ∀q, where e is a variable q'.

[9] In our original (A9) construed as an axiom of Q, 'closure' was intended as first-order closure where a first-order closure of $A \in \mathrm{WFF}_Q$ is $\ulcorner \forall x_1 \ldots \forall x_n(A)\urcorner \in \mathrm{WFF}_Q$, where $\langle x_1, \ldots, x_n\rangle$ is an arbitrary permutation of the free individual variables in A. In (A9), rewritten for Q+, and in (A22) let 'closure' be understood as what might be termed 'general closure', where a general closure of $A \in \mathrm{WFF}_{Q+}$ is $\ulcorner \forall \alpha_1 \ldots \forall \alpha_k(A)\urcorner \in \mathrm{wff}_{Q+}$, where $\langle \alpha_1, \ldots, \alpha_k\rangle$ is an arbitrary permutation of the free individual and propositional variables in A. This is a point important for the relative consistency proof I outline for Q+ in Section 4, and I am grateful to Evan W. Conyers for calling it to my attention.

[10] See, for example, Boolos and Jeffrey (1982, Chap. 14).

[11] Here and throughout I follow Boolos and Jeffrey's format for defining representation. See Boolos and Jeffrey (1982, p. 158).

[12] For any number n which is not the Gödel number of a c-formula we can let $f'(n) = \nu$, where ν is chosen so as not to be the Gödel number of any c-formula. This will obviate any need to deal with partial functions.

[13] See Boolos and Jeffrey (1982, p. 174).

[14] Here I use $c_i, c_{ii}, c_{iii}, \ldots, c_r$ in order to avoid confusion with $c_1, c_2, c_3, \ldots, c_n$, as employed in outlines of representation elsewhere in the paper. Throughout the paper c_1, $c_2, c_3, \ldots, c_n$ are used as variables for any c-formula. $c_i, c_{ii}, c_{iii}, \ldots, c_r$, in contrast, are to be envisaged as particular c-formulae indexed by their order of appearance in a given enumeration. In the following few paragraphs I use j and k as index variables. c_j and c_k are thus to be thought of as the *jth* and *kth* c-formulae of our enumeration, respectively.

As a formula, of course, any closed formula c_j of our enumeration of c-formulae will also appear as some f_m in our enumeration of *all* formulae. In fact, such an m will be effectively determinable from j; m will equal g(j) for some recursive function g.

[15] Here and throughout I have confined myself to closed formulae as counter- or encoding formulae. Use of closed and open formulae of a particular level is perhaps technically possible as well, but leads to at least interpretational complications. I have chosen to avoid these difficulties here.

[16] We can treat *d* as a total function by the mechanism of Note 12.

[17] Relative consistency proofs with this basic form but for simpler and more familiar systems appear in Church (1956) and Hunter (1971). Similar proofs are also possible for systems richer than Q+ in various ways.

[18] It would not suffice to replace '$\triangle(e_I)$' with simply 'e_I' here, since '$\triangle(p)$' is a wff of $Q + \triangle$ but 'p' is not a wff of $Q+'$.

[19] See, for example, Boolos and Jeffrey (1982, p. 173).
[20] Robert F. Barnes has suggested a proof for this form of the diagonal lemma assuming representation not of a c-diagonalization function of c-formulae $diag^c$ but of an instantiation function of c-formulae *INST*, thereby avoiding formulae with free variables. Axiomatic machinery for representation of *INST* can be introduced by essentially the same strategy outlined in Section 2.

Suppose ⊠ represents *INST*:

$\vdash_T \forall p(\boxtimes(c_e, c_g, p) \equiv p = c_h)$
whenever f_h is the result of instantiating the outer propositional quantifier of f_e with c_g.

Let f_j be: $\forall q \forall p(\boxtimes(q, q, p) \rightarrow B(p))$.
Let f_k be: $\forall p(\boxtimes(c_j, c_j, p) \rightarrow B(p))$
Then:

$\vdash_T f_k \rightarrow (\boxtimes(c_j, c_j, c_k) \rightarrow B(c_k))$
$\vdash_T \boxtimes(c_j, c_j, c_k) \rightarrow (f_k \rightarrow B(c_k))$
$\vdash_T \forall p(p = c_k \rightarrow \boxtimes(c_j, c_j, p))$
$\vdash_T c_k = c_k \rightarrow \boxtimes(c_j, c_j, c_k)$
$\vdash_T c_k = c_k$
$\vdash_T \boxtimes(c_j, c_j, c_k)$
$\vdash_T f_k \rightarrow B(c_k)$ (*)
$\vdash_T \forall p(\boxtimes(c_j, c_j, p) \rightarrow p = c_k)$
$\vdash_T \forall p(p = c_k \rightarrow ((B(c_k) \rightarrow B(p)))$
$\vdash_T \forall p(\boxtimes(c_j, c_j, p) \rightarrow (B(c_k) \rightarrow B(p)))$
$\vdash_T \forall p(B(c_k) \rightarrow (\boxtimes(c_j, c_j, p) \rightarrow B(p)))$
$\vdash_T B(c_k) \rightarrow \forall p(\boxtimes(c_j, c_j, p) \rightarrow B(p))$, (**)
i.e., $\vdash_T B(c_k) \rightarrow f_k$
$\vdash_T f_k \equiv B(c_k)$ by (*), (**).

[21] Here assumed schemata are in fact of a form weaker than those standardly introduced for the Knower, which would demand for (2) not simply $\triangle(\bar{\bar{T}})$ for some convenient tautology T but the more complex and embedded $\triangle(\overline{\overline{\triangle(\bar{\bar{A}}) \rightarrow A}})$ in particular (see for example Anderson (1983), Kaplan and Montague ([1960]/1974), and Grim (1991)). One lesson of the form of the proof below, due to Robert F. Barnes, is that this stronger assumption is not strictly necessary.
[22] See also Note 15.
[23] On the assumption that these are consistent, of course.

With regard to consistently adding schemata for the Knower to a system such as Q, Montague appeals to "perfectly natural model-theoretic interpretations" (Montague [1963]/1974, p. 294). Evan W. Conyers has shown me an elegantly simple syntactical proof that if Q is consistent, so is an extension Qs with the standard Knower schemata. Conyers's proof uses an 'eraser function' which systematically purges formulae of the operator $\triangle$.

A similar eraser function will translate Q + s, an extension of Q+ in which these schemata appear restricted to closed formulae A and B, into Q+. We have shown that Q+ is consistent if Q is, so Q + s will be consistent if Q is.
[24] It is tempting to read '$\triangle(\bar{\bar{A}})$' in the first schemata as 'the formula with counter-formula

$\bar{\bar{A}}$ is necessary', for example, and to read 'Δ(A)' in the second set as 'formula A is necessary', making it appear that the only difference between the two cases is how formula A happens to be referred to.

REFERENCES

Anderson, Alan Ross and Nuel D. Belnap, Jr.: 1975, *Entailment*, Vol. I, Princeton University Press, Princeton.

Anderson, C. Anthony: 1983, 'The Paradox of the Knower', *Journal of Philosophy* **80**, 338–55.

Asher, Nicholas and Hans Kamp: 1989, 'Self-Reference, Attitudes and Paradox', in Gennaro Chierchia, Barbara H. Partee, and Raymond Turner (eds.), *Properties, Types and Meaning*, Vol. I, Kluwer Academic, Dordrecht, pp. 85–158.

Boolos, George S. and Richard C. Jeffrey: 1982, *Computability and Logic*, 2nd ed., Cambridge University Press, New York.

Bull, R. A.: 1969, 'On Modal Logic with Propositional Quantifiers', *Journal of Symbolic Logic* **34**, 257–63.

Burge, Tyler: 1978, 'Buridan and Epistemic Paradox', *Philosophical Studies* **34**, 21–35.

Burge, Tyler: 1984, 'Epistemic Paradox', *Journal of Philosophy* **81**, 5–28.

Church, Alonzo: 1956, *Introduction to Mathematical Logic*, Princeton University Press, Princeton.

Fine, Kit: 1970, 'Propositional Quantifiers in Modal Logic', *Theoria* **36**, 336–46.

Fine, Kit: 1977, Technical Appendix to A. N. Prior and Kit Fine, *Worlds, Times, and Selves*, Duckworth, London, pp. 162–68.

Grim, Patrick: 1988, 'Truth, Omniscience, and the Knower', *Philosophical Studies* **54**, 9–41.

Grim, Patrick: 1991, *The Incomplete Universe: Totality, Knowledge, and Truth*, MIT Press/Bradford Books, Cambridge.

Hunter, Geoffrey: 1971, *Metalogic*, University of California Press, Berkeley.

Kaplan, David: 1970, Abstract of 'S5 with Quantifiable Propositional Variables', *Journal of Symbolic Logic* **35**, 355.

Kaplan, David and Richard Montague: [1960]/1974, 'A Paradox Regained', *Notre Dame Journal of Formal Logic* **1**, 79–90 (reprinted: 1974, in Richmond H. Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven, pp. 271–85).

Montague, Richard: [1963]/1974, 'Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica*, Vol. 16, Societas Philosophica Fennica, Helsinki, pp. 153–67 (reprinted: 1974, in Richmond H. Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven, pp. 286–302).

Putnam, Hilary: [1967]/1983, 'Mathematics Without Foundations', *Journal of Philosophy* **64**, 5–22 (reprinted: 1983, in Paul Benacerraf and Hilary Putnam (eds.), *Philosophy of Mathematics*, 2nd ed., Cambridge University Press, Cambridge, pp. 295–311).

Reinhardt, William N.: 1980, 'Necessity Predicates and Operators', *Journal of Philosophical Logic* **9**, 437–50.

Riviéres, J. des and H. Levesque: 1986, 'The Consistency of Syntactical Treatments of

Modality', in Joseph Y. Halpern (ed.), *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the 1986 Conference*, Morgan Kaufman, Los Angeles, pp. 115–30.

Skyrms, Brian: 1978, 'An Immaculate Conception of Modality', *Journal of Philosophy* **75**, 368–87.

Thomason, Richmond H.: 1977, 'Indirect Discourse is Not Quotational', *Monist* **60**, 340–54.

Group for Logic and Formal Semantics
Department of Philosophy
State University of New York
Stony Brook, NY 11794–3750
U.S.A.