To: libstx@osu.edu
From: OSU Interlibrary Services <liblend@osu.edu>
Subject: DOCUMENT DELIVERY REQUEST 458774

DOCUMENT DELIVERY for Neil Tennant

ILLiad Transaction Number: 458774

Journal Title: Dialectica
Journal Vol: 36
Journal Issue:
Journal Year: 1982
Article Pages: 265-296

Article Author: Tennant.Neil
Article Title: Proof and Paradox

Notes: This is my own article, of which I badly need a soft copy. It was written before the age of the personal computer!

If you need to contact us, please refer to this Document Delivery Transaction Number: 458774
E-Mail: liblend@osu.edu
Phone: 614-292-6211

NOTE TO SCANNING LIBRARY:
Send this Document Delivery article to patron via Odyssey. ILLiad locations should 'Mark Item as Not Found' if requested article is not available.

80201057̶06003

Proof and Paradox

by Neil Tennant*

> There is a great deal of talk about [paradoxes] in the mathematical literature. But you may hardly conclude from this that there exists a general agreement about the sense of the word. To my mind it is much more likely that most who use it don't know exactly what sense they attach to it. Whether there is someone who has found a tenable explanation of the word, I don't know. Many do not explain the word at all, which is the most convenient thing to do, but would only be justifiable if there was a general agreement among people about it.**

Introduction

This paper investigates the logico-semantical phenomena loosely and collectively known as the paradoxes. The approach will be essentially different from any other known to me in the literature. For I shall be seeking a *proof-theoretic* characterisation of the items in question, items that appear always to be subsumed into the domain of semantics or interpretation. It seems to me, however, that a proof-theoretic approach can yield useful insights into the structural features at issue; and can even, if developed far enough, pose an interesting 'completeness problem' for the paradoxes parallel to that for logical truth and consequence in semantically open languages.

* University of Stirling, Department of Philosophy, Stirling FK9 4LA, Scotland
** Apologies to Frege in *Posthumous Writings* (trans. P. Long & R. White), 'Logical Defects in Mathematics', p 159.

How does this choice of problem and approach fit into the study of rationality, or contribute to our understanding of it? Presumably one wishes to command as clear a view as possible of various kinds of reasoning, from practical reasoning in the theory of action, to logical inference in the foundation and rigorous reconstruction of scientific theories. A theory of rational thought processes might be both descriptive and normative. Meaning is rooted in actual usage. Normative laws of thought, insofar as they attain their normative character by faithfully reflecting diagnosable meaning, are constrained by descriptions of both the vehicles and actual motions of thought. The vehicles are the representations, both mental and linguistic, which the rational mind processes; and their motions depend both on their own inherent design and on the thrusts they receive from an imaginative intellect. I would argue moreover – though this is not the place to do so – that all that intelligence, grasp of meaning and evaluation consists in 'internally' is the activity, or ability to engage in the activity, of transforming structured representations of finite size. Among such representations are included the most peripheral, or 'earliest' perceptual constructs transmitted from sensory surfaces, as well as the 'latest' neural patterns governing motor responses. What happens 'in between' is the internal aspect just referred to. The transformations in question need not be conscious, nor even accessible to introspection. The internal or mental representations could be further coded by external symbolic means – such as in written proofs, or sentence utterances – which themselves could in turn be the source of further perceptual constructs. The latter would have to be suitably related, in a functionalist way, to the mental representations involved in the grasp or formulation of arguments or meanings not yet coded in an external linguistic or symbolic form. This general sort of claim is a working hypothesis in the field of artifical intelligence. It seems forced upon one by any materialist or functionalist theory of mind. Note that it does not amount to a mere Turing test of intelligence. It is quite compatible with the view — which I hold — that attributions of intelligence or rationality can only be made on the basis of gross observable behaviour. Minds must be embodied, and those bodies exhibit complex agency before one can concede that even complicated looking symbolic products – such as 'English' on a console screen – should be interpreted in the way Turing assumed it could.

All this may seem far removed from the central concern of this paper; but this is not the case at all. For I am concerned to tackle the problem of the paradoxes, which have always been a great stumbling block to a fully comprehensive account of inference within suitably rich languages. The richness in question is, of course, semantic closure. It consists in a certain apparatus of reference and of predication that precipitates paradox. If only we can understand *how* this happens we shall be in a better position to decide *what to do* about it in the so-called rational reconstruction of science. Perhaps the rationality of this enterprise depends not so much on purging, as on devising ways of accomodating paradox. This, however, already takes me beyond the project at hand. That project is to understand *how* paradox arises and what light can be shed on it by a theory concerned specifically with transformations of structured representations – in this case, with transformations of *proofs*.

It would not be too harsh a caricature to say that our present conception of the paradoxes (with the possible exception of Martin and Woodruff's, and Kripke's seminal accounts) is about as pre-systematic and imprecise as were our conceptions of *modality* and *opacity* before the rise of intensional semantics of various kinds; or even of *logical truth* and *logical consequence* before the Bolzano-Frege tradition. The beginner's general question "What is a paradox?" usually receives the following sort of answer. First one cites a paradigm case, such as the Liar paradox. Then one goes on to mention the impossibility of *any* truth value assignment to the sentence in question. Finally one gestures towards self-reference, cyclic reference, semantic closure, vicious definitional circularity, or even inconsistent pre-systematic beliefs, expectations and intuitions as the source of the trouble. The fortunate beginner will then be put onto a list of reading on the well-known 'diagnoses' and proposals for blocking or reform – language hierarchies, stratification of formulae, indexical truth, grounded truth-value assignments, third truth values, truth value gaps, logics of paradox and so on.

Let us replace our question with a more tractable one: "What is a logical truth?" It usually receives the following sort of answer. First one cites a paradigm case, such as the law of contradiction. Then one goes on to mention the impossibility of *any* falsifying truth value assignment to the sentence in question. Finally one gestures towards the meanings of the logical operators, as codified in truth tables or truth-definitional clauses, and towards the range of possible interpretations, as represented by rows in a truth table or by extensions in a model, as the source of the property in question. This is the first step in explicating logical truth as truth in all possible worlds.

But what is very different in this case is that the initial answer about logical truth can be amplified and refined. Logical truth can be defined in a precise fashion for semantically open languages by semantical theories drawing on set theory for their formulation. The defined property of logical truth is, however, ideal and infinitistic in character, involving as it does quantification over all models, which are themselves defined as having (arbitrary) sets as domains, and as assigning to each non-logical expression of the language an appropriate extension within an appropriate Cartesian power of the domain. It is thus apparent that we should in general experience some difficulty in identifying logical truths, or as recognising them as such within the language concerned. Thus even before Church's theorem showed that the latter was indeed a recursively unsolvable problem, there was a genuine and legitimate demand for some reduction of the notion of logical truth, as would be provided by a suitable generative axiomatic system. What was in effect required was a class of checkable constructions – that is, proofs – that would establish just the logical truths as such. Thus proof systems and the completeness problem were born, and Gödel's completeness theorem for first order logic provided the reduction sought.

In the light of this analogy, let us now return to the case of the paradoxes. Some modern authors have begun to take the first steps in a semantic characterisation of the paradoxes (and I shall be referring especially to Kripke's work in what follows). It is to be hoped that we shall thereby amplify and refine our notion of paradoxicality to the point where analogous problems can be posed and solved. The famous paradoxes, such as the Liar, are very simple. They are thus like the law of excluded middle, or the law of contradiction, and similar tautological trivia from which a richer notion of logical truth is extracted. If we can eventually extract a correspondingly rich notion of paradoxicality from the semantical accounts mentioned, there might then arise a genuine and legitimate demand for some reduction of this notion. The resulting proof-like constructions might then turn out to be as important a tool for epistemology as the conventional notion of proof, should semantically closed languages (say) ever become acceptable for certain scientific purposes.

In this paper I accordingly motivate some proof-theoretical considerations in order to formulate a completeness conjecture for the paradoxes. The reader will note that I have not yet spoken of 'diagnosing' the paradoxes.

Nor shall I be advocating any therapy or reform. I am content to play the pathologist, not the physician. So my terminology will be neutral – "identifying", "characterising", "defining", "giving criteria for" and so on; and, as we finally locate the 'problem' in a more mathematical setting, even: "axiomatising".

But for those who think in terms of diagnosis and remedy, who share the Tarskian sentiment that "The appearance of an antinomy is . . . a symptom of disease", I shall offer a mild palliative. I shall always try to present my proof-theoretic considerations within *intuitionistic* logic. This should at least

allay the suspicion that bivalence or excluded middle or some jaundiced relative is the source of contagion. (Frege perhaps ranks as the first notable to have wondered whether this might indeed be the case.)

1. Some well-known paradoxes in a proof-theoretic setting

In this section, and in section 5, I shall isolate an important property of proofs in connection with paradoxes in the Liar family, Grelling's paradox, Curry's paradox and Tarski's quotational paradox.

Each of the proofs to be considered is a proof of absurdity (Λ) involving the paradoxical sentence (or sentences) in question. But of course paradoxicality is not the 'mere inconsistency' of these as might be established by such a proof. To take a trivial example, the proof

$$\dfrac{\dfrac{A \,\&\, —A}{A} \qquad \dfrac{A \,\&\, —A}{—A}}{\Lambda}$$

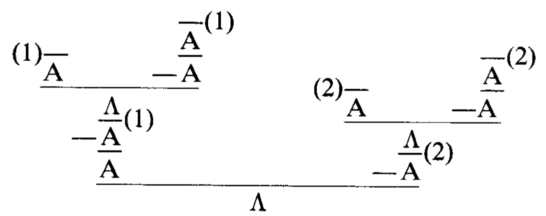
shows that A & —A is inconsistent. But does it tell us whether A & —A is paradoxical? Of course it does! One can see from the proof that A & —A is *not* paradoxical, and only inconsistent. For it is a proof of Λ from A & —A *in normal form.*

The reader may now ask "But then isn't the notion of paradoxicality vacuous? Does not every paradox imply a contradiction, and does not every such proof reduce to one in normal form?" The answer to the first question is negative. While admittedly every paradox is associated with a proof of Λ, *the proofs concerned do not reduce to ones in normal form.* This is because the reasoning takes place in systems for semantically closed languages. In such systems normalisation theorems cannot hold.

We have above a paradigm 'bare faced' proof of Λ from A & —A. Now let us give a similarly simple example of the problem feature of proofs connected with paradoxes. The reader will no doubt be familiar with the way in which much of the reasoning in connection with the various paradoxes has an "if and only if not" flavour. Let us represent this schematically by means of the two rules of inference

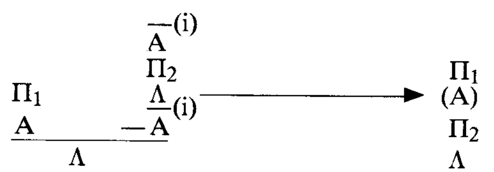
$$\dfrac{A}{—A} \qquad \dfrac{—A}{A}$$

which we shall not for the time being "unpack" in particular cases in pursuit of more detailed legitimation. Now consider the following simple proof of Λ:

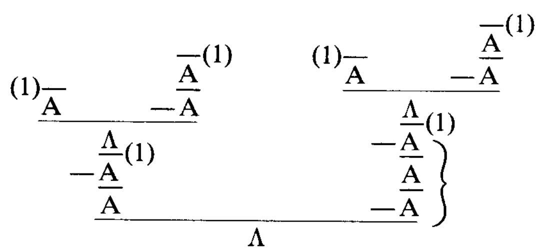


(Note that this proof appears to establish Λ from no assumptions at all — as though the very language itself were incoherent, rather than any particular claim A formulable within it.)

The reader will note that the bottommost occurrence of —A is *maximal* — it stands as the conclusion of negation introduction and as the major premiss of negation elimination. We can therefore attempt to remove it by means of the reduction procedure



We then obtain



If we now replace the prolix portion indicated (from —A to —A via A) by —A, we obtain our original proof. I shall always assume that eliminating detours such as the one mentioned, that lead from a sentence occurrence to another occurrence of the same sentence, is an acceptable part of any procedure of normalising proofs. The normalisation sequence thus enters a loop, never terminating with a proof in normal form.

It is clear that paradoxicality hinges partly on the nature of the inferences from A to —A and from —A to A. The proof above lays this bare as simply as possible. But not every paradox need display this feature so clearly. It is therefore of considerable interest to enquire after techniques for discerning, in less direct fashion, whether something at root similar to this circularity of inference is at work in all paradoxes. I wish to maintain that it is indeed their distinguishing feature. I propose precisely the test of non-terminating reduction sequences. We shall in due course examine more closely some well known paradoxes with this test in mind.

They are all formulated in a language permitting reference to its own expressions. There are simple ways of disinguishing use and mention but it will always be clear from context, without additional notation such as corner quotes, which sense I intend. Thus where T is a truth predicate, say, I shall write $\alpha \dashv\vdash T\alpha$ instead of $\alpha \dashv\vdash T\ulcorner\alpha\urcorner$.

(In theories containing sufficiently strong arithmetic, α is often defined as the numeral for the code number of the expression a. But then 'self-reference' is really illusory, since the reference is still to the *numbers* concerned. It is only indirectly, through the coding and representability of syntactical relations in arithmetic, that one achieves the apparent 'self-reference' of the unprovable Gödelian sentence, for example. My own notational convention — albeit often suppressed in practice — would cater for all genuine cases of reference to expressions.)

For the *Liar paradox* to arise the language must first contain a truth predicate T such that

for all α $\alpha \dashv\vdash T\alpha$

From left to right this is like a rule of T-introduction, while from right to left it is like a rule of T-elimination. Furthermore, the language must contain some particular sentence λ (the Liar) 'interdeducible' with its own falsity predication:

$\lambda \dashv\|\vdash —T\lambda$

I say 'interdeducible' in scare quotes, because the inferences in question are what I would call *id est* inferences. Referentially, given the structure of λ, things so turn out that λ *is* none other than the sentence —Tλ. I shall use the new turnstile above to express this, and in proofs I shall label the corresponding inferences "ie". The following is the proof of Λ associated with the Liar, with its looping reduction sequence:

SEE DIAGRAM 1

Note that it is necessary, in order for paradox to arise, that we have some particular sentence (or set of sentences) like λ for which an 'interdeducibility' condition like the one above holds. (Such will be found to be the case also in the cyclically referential cases, or the post card paradox, which are attenuated and more leisurely versions of the Liar, producing paradoxical proofs of Λ only after a longer chain or tree of inferences.) The problematic sentences are essential. It is not enough that the language contain a truth predicate. I owe to Graham Priest the observation that in first order logic with Hilbert's epsilon operator, for example, one can consistently have a truth predicate.[1]

If $—\epsilon x(x=x) = \epsilon x(—x=x)$ is an axiom, and if we define $\ulcorner\alpha\urcorner$ as $\epsilon x\alpha$ (x not in α), and Tx as $x=\epsilon y(y=y)$, then the truth schema above is derivable.

In the language of the Liar, semantic closure is forcefully apparent on the referential side, and not just on the side of semantic predication. Sentences may contain their own names, or at least be interdeducible with others containing those names. But the assumption of a truth predicate is weaker than the assumption of a satisfaction predicate. Let us now consider a language with the latter. In it the following holds:

$$\text{for all } \alpha \ S(t, \alpha) \dashv\vdash \alpha^x_t$$

Now consider, on the referential side, a weaker condition than before: this time merely to the effect that each formula has a name. Now consider in place of α the formula —S(x, x) ("x is heterological"), and in place of t the term $\ulcorner$—S(x, x)$\urcorner$. *Grelling's paradox*[2] has the following looping reduction sequence of proofs:

SEE DIAGRAM 2

In both the foregoing examples negation featured prominently. It is therefore worth noting that negation is not an essential feature or ingredient of paradoxicality. It is simply not true that paradox arises just in case a sentence is interdeducible with its own negation. For consider *Curry's paradox*. For this, as with the Liar, we assume a truth predicate satisfying the usual schema, but now, in addition, for any given p we assume some particular γ (the p-Curry) 'interdeducible' with is own "true only if p"-predication:

$$\gamma \dashv\vdash T\gamma \supset p$$

(Note that via representability and diagonalisation in sufficiently strong arithmetic we have the fixed point theorem

[1] G Priest, "The Logical Paradoxes — a unified account". Paper delivered to the Australasian Association of Logic, Melbourne University 1977.

[2] This formulation is due to Priest, *loc cit.*

For any predicate ψ there exists a sentence β such that

$$\beta \dashv\vdash \psi\beta$$

which suffices to yield interdeducibility schemata like those required for the Liar and the Curry paradoxes. But of course we can then use the 'paradoxical' arguments to show that arithmetic can't contain a truth predicate. The fixed point inferences in arithmetic, it is worth noting, are not *id est* inferences. There is a lot of unpacking to do, via representability. And of course the argument will normalise, since it is a proof in first order, semantically open language.) The looping reduction sequence of proofs in the case of the Curry paradox is as follows:

SEE DIAGRAM 3

The reader can verify that the same phenomenon will be observed if the proof is re-cast in the (&, —)-fragment, with $t\gamma \supset p$ replaced by $—(T\gamma \,\&\, —p)$.

Let us now move on to a more complicated example, *Tarski's quotational paradox*. For this we need what I would call the disquotational Leibniz law

$$\frac{\alpha \quad \ulcorner\alpha\urcorner = \ulcorner\beta\urcorner}{\beta}$$

as well as the assumption that, for each term α for sentences, we can form the sentence

$$\forall p(\alpha = \ulcorner p\urcorner \supset —p)$$

(which for convenience I shall call γ) involving quantification into quotes. As before, a proof of Λ involving γ has a looping reduction sequence:

SEE DIAGRAM 4

By now the reader should be willing to believe that something general is going on. Paradox assumes the form of a proof of Λ or of an arbitrary p by

(i) using a general interdeducibility schema (such as for truth or satisfaction), and

(ii) exploiting the presence of some sentence (perhaps depending on p) that gives the reasoning its particular flavour.

But neither this sentence nor any other (but see below for qualification of this claim) appears as an undischarged assumption in the proof of Λ or of p. The proof has no undischarged assumptions (except perhaps those that make empirical claims that precipitate paradox in sentences that are not intrinsically paradoxical — again, see below). Moreover the proof cannot be normalised. (Later we shall see that perhaps only the 'pure' paradoxes have proofs of Λ or of p with no undischarged assumptions.)

In a certain sense the proof(s) for the Liar paradox essentially involve the Liar sentence λ, namely —Tλ; and those for the Grelling paradox essentially involve the Grelling sentence —S(t, t), and so on. This is what makes a proof one "associated with" or "connected with" each of these paradoxes, respectively. Now just as in the standard case, where we need and have an effective method for determining the class of undischarged assumptions and the conclusion of any given proof, so we need and have a method for determining the set of paradoxical sentences with which one of these non-normalisable proofs is associated. One simply forms the set of sentences involved in the *id est* inferences within the proof.

In some cases the proof may be highly abnormal, susceptible to a considerable amount of normalisation before the reduction sequence enters a loop. In such a proof we may inadvertently collect more sentences into the set just described than are absolutely necessary for the paradox to arise. This is just like the case in standard logic where we inadvertently count among the assumptions of a proof some that are not absolutely necessary (even in the context of *that* proof) in order to secure the conclusion. Such assumptions are those that disappear from the scene when the proof is normalised. (Precisely this is the main idea behind my treatment of the problem of entailment — see 'A Proof-theoretic Approach to Entailment', *Journal of Philosophical Logic* 9 (1980) 185–209.)

Similarly in the case of paradoxical proofs we can at least normalise them to some point Π in the reduction sequence at which the ensuing loop re-enters, so to speak. Then we can read off the sentences involved at Π, in order to obtain the most economical identification of the paradox in question.

2. *Russell's Paradox*

Russell's paradox is usually claimed to arise within naive set theory. But in the spirit of the inferential approach above, I would like to suggest that we regard it as arising within the naive *logic* of sets. Consider such a naive logic with the following two inference rules Πϵ (for "predication-membership") and ϵΠ (for "membership-predication"):

$$\Pi\epsilon \quad \frac{\psi^x_t}{t\epsilon\{x|\psi\}} \qquad\qquad \epsilon\Pi \quad \frac{t\epsilon\{x|\psi\}}{\psi^x_t}$$

Πϵ might be regarded as the rule of introduction for the set term forming operator, and ϵΠ as the corresponding rule of elimination. (This is indeed how Prawitz presented the matter in his monograph *Natural Deduction — A Proof-Theoretical Study*.) Apart from these two new rules, the naive logic of sets is based on the implicit background assumption that the language is logically perfect in Frege's sense — that is, every term of the language denotes. Now if we let ϱ be the term $\{x|—x\epsilon x\}$ we obtain the following proof with the looping reduction sequence shown:

SEE DIAGRAM 5

By our criterion, this would seem to indicate that we have a genuine case of paradox.

It may be thought that the basic mistake here was to have assumed a logically perfect language in the sense above. Would not the appearance of paradox be dispelled if we used instead a *free* logic, and modified the rule Πϵ above to

$$\Pi\epsilon' \quad \frac{\psi^x_t \quad \exists x\, x=t \quad \exists y\, y=\{x|\psi\}}{t\epsilon\{x|\psi\}} \quad ?$$

The answer is negative. We still obtain a looping reduction sequence:

SEE DIAGRAM 6

This does not yet show, however, that the fault does not lie in our choice of rules. At this point I would advocate the logic of sets in my book *Natural Logic*, § 7.10, which differs from the naive logic discussed here in the following crucial respect. In this logic the rules for introducing and eliminating the set term forming operator are formulated for contexts of *identity* rather than membership, as befits term-forming operators in general. That is, the introduction rule tells us the conditions under which we are entitled to infer the *identity* claim $t=\{x|\psi\}$, and the elimination rule spells out the consequences or commitments of such a claim. The rules are as follows:

$$\{\}\text{–I} \quad \frac{\begin{matrix}\overline{a\epsilon t}^{(i)} \\ \vdots \\ \psi^x_a\end{matrix} \qquad \exists!t \qquad \begin{matrix}\overline{\psi^x_a}^{(i)} \quad \overline{\exists!a}^{(i)} \\ \vdots \\ a\epsilon t\end{matrix}}{t=\{x|\psi\}}\ (i)$$

where a does not occur in $t=\{x|\psi\}$ nor in any undischarged assumptions of the subordinate proofs other than those of the form displayed

$$\{\}\text{–E1} \quad \frac{t=\{x|\psi\} \quad u\epsilon t}{\psi^x_u} \qquad\qquad \{\}\text{–E2} \quad \frac{t=\{x|\psi\} \quad \psi^x_u \quad \exists!u}{u\epsilon t}$$

(Note also that this logic is free, as shown by the existential premisses variously required. This means that there are two other rules that might be unfamiliar to the reader. Reflexivity of identity is expressed by the rule

$$\text{R}\quad \frac{\exists !\, t}{t = t}$$

and there is a denotation rule for atomic sentences (including identities) involving t not within the scope of any occurence of the term-forming operator:

$$\text{D}\quad \frac{A}{\exists !\, t}$$

This rule captures the semantical requirement that a term's failure to denote renders an atomic predication involving it false.)

The naive logic of sets provided introduction and elimination rules that added up to Church's conversion schema "$t\epsilon\{x|\psi\}$ iff ψ^x_t". This schema, with suitable existential qualifications for freedom of t, now comes out in each direction as a special case of E1 or E2.

The general claim I am advancing is that one ought to get the introduction and elimination rules right on philosophical grounds — from considerations of meaning and reference —before pronouncing on the 'paradoxical' status of certain sentences involved in proofs proceeding in accordance with those rules. But if we re-formulate the Russell proof in the free logic of sets just given, we obtain the following:

SEE DIAGRAM 7

This is an intuitionistic proof in normal form, reducing the existential assumption $\exists !\varrho$ to absurdity. But we can still discern a certain prolixity in the final proof. If the occurrence of $\sim \varrho\epsilon\varrho$ therein (marked by $\dagger$) is identified with the similar one in Π, thereby eliminating the prolixity in the 'final' proof, one obtains once again a maximal occurrence of $\sim \varrho\epsilon\varrho$ and enters a reduction loop. (I owe this observation to Alan Weir). It thus appears that even with the best philosophical will in the world, Russell's remains an intrinsically troublesome case of paradox. So long as one wishes to have the set term forming operator {} primitive, even within our sophisticated free logic of sets the proof reducing $\exists !\varrho$ to absurdity is prolix.

Similar observations hold for the *'Curry set' paradox*. The following is an improved formulation of the one given by Meyer, Routley and Dunn. For arbitrary q, consider the set abstract $C =_{df} \{x | x\epsilon x \supset q\}$. Let p be the sentence $C\epsilon C$. Then the naive inference schemata $\Pi\epsilon$ and $\epsilon\Pi$ have as special instances

$$\frac{p \supset q}{p} \qquad\qquad \frac{p}{p \supset q}$$

respectively. Using these we construct the proof

SEE DIAGRAM 8

As the reader can verify, this proof has a looping reduction sequence. If, however, we re-formulate it in the free logic of sets suggested above, this is no longer the case. For then we have the proof

$$\begin{array}{c}\exists !C\\ \Pi\\ C\epsilon C\end{array}\ :$$

SEE DIAGRAM 9

and using it we can construct the normal proof

SEE DIAGRAM 10

of q from the existential assumption $\exists !C$. So for refutable q the claim "$\{x|x\epsilon x \supset q\}$ does not exist" is now provable. But a similar prolixity infects the 'final' proof, as remarks similar to those made above in the case of Russell's paradox would show. This seems to discourage the hope that there may be a proof-based distinction between the set theoretical paradoxes (as arising out of philosophical muddle, from confusion about meaning) and the logico-semantical paradoxes.

But at least in the set-theoretical case we can avoid even the residual prolixity of the proof reducing $\exists !\varrho$ to absurdity by choosing to treat the set term forming operator {} as defined, and not primitive.[3] A similar strategy is not, however, available in the case of the logico-semantical paradoxes. I for one can see no objection to the rule

$$\frac{\alpha}{T\alpha}$$

balanced by its elimination counterpart $\frac{T\alpha}{\alpha}$. Nor can I formulate any principles of meaning and reference which would guarantee either that not every expression can be referred to, or that it is impossible for a term to denote any sentence in which it occurs or which is interdeducible with one in which it occurs. (cf Kripke, 'Outline of a Theory of Truth', p 693).

If we accept semantic closure in any of these forms, and are confident that the rule(s) $\alpha \dashv\vdash T\alpha$, say, capture the content of T in a satisfactory way, then we have to decide upon a rational accommodation of the logico-semantical paradoxes.

[3] cf Takeuti & Zaring

3. *On rational response to the paradoxes*

I fully realise how inadequate any supposedly final word on this matter would be. In this section I shall give only the briefest list of responses one might make in the face of paradox.

Firstly, one could take the view that, like any other industry logic has its toxic by-products. But so long as these can be sealed off and contained, there is no neeed for alarm. I shall call this the 'waste disposal' theory. Appropriate equipment to serve it might be an effective method for identifying paradoxes, plus a paraconsistent logic. The method is that of inquest and safe burial.

Secondly, one could take the view that paradoxes are linguistic creatures with benign recessive genes in homozygous condition. Alleles for self-reference and for semantic predication stay in the linguistic gene pool because of superior homozygote fitness. Ingredients that sometimes combine to produce paradox can yield compensating benefits in isolation. I shall call this the Darwinian theory of paradox, and venture to say that it might not be altogether as bizarre as the reader no doubt thinks on the basis of this crude sketch. Why should not linguistic evolution produce its own unviable mutants the way organic evolution does? This theory might be served by a technique for listing the paradoxes, so that appropriate eugenic measures can be taken to weed out the mutants.

Then there is the 'deterrence' theory. Historically it is the oldest. The deep silos of the language hierarchy, the thick concrete of stratification, the SALT agreements ("Sets are little things") are all made to contain the problem. But do they provide ultimate safeguards? *Is* ZF consistent?

Finally, one might refuse to take part in the arms race of semantic ascent. One can remain ontologically neutral, even nominalist, retreating to showing instead of saying.

I suggest that we could be helped to make the best policy decision by metalogical results about the possibility of detecting, or screening for paradox. That is, I propose an investigation of the decision and completeness problems for paradoxicality in semantically closed languages via a proof-theoretic reduction of their semantical definition.

But first let us test our proposed criterion for paradoxicality on one new case. It is a case that I discovered *after* formulating the criterion, and for which my pre-formal intuitions (*pace* the author of this alleged paradox) were in favour of the 'mere inconsistency' view. The test confirmed these intuitions. I refer here to Chihara's *secretary liberation 'paradox'* (which he actually attributes to Cioffi).

4. *The 'Paradox' of Sec Lib*

Thus Chihara:

> Consider now a *full-fledged semantic* paradox. Imagine a situation in which many clubs have hired secretaries but have established rules excluding such secretaries from membership. Suppose that these secretaries form their own club, Secretary Liberation (or "Sec Lib" for short), the rules of which state: "A person is eligible to join this club if, and only if, he (she) is secretary of a club which he (she) is not eligible to join." All goes well for the club until it hires itself a secretary, a certain Ms Fineline, who has the misfortune of being a secretary of no other club. The paradox arises: Is she, or is she not, eligible to join Sec Lib? On the assumption that she is, it follows that she is not; and if she is not, she is[4].
>
> The contradiction is derived from:
>
> (1) For every person x, x is eligible to join Sec Lib iff there is a club of which x is a secretary and which x is not eligible to join.
>
> (2) Ms Fineline is secretary of Sec Lib.
>
> (3) Ms Fineline is not secretary of any other club.
>
> (2) and (3) are just given facts that are empirically determinable. So this suggests that (1) is false. But why is one inclined to think that (1) is true? . . . it would seem that (1) has been made true by fiat: after all, *that is what the rules say* . . . In this case, it is thought that one can make a statement that is inconsistent with statements of fact true by fiat. But one can no more do (this) than one can (make an inconsistent pair of statements true by fiat). Hence, we should reject (1).
>
> But there are special reasons why most people do not think of questioning (1). It is hard to question the premise since the eligibility rules *seem to be in order,* as can be seen from the fact that they work in general: in most situations, Sec Lib's rules function without difficulty, leaving no doubt as to whether or not a candidate is eligible. The idea that a general rule might be perfectly adequate in most situations and yet defective, or even inconsistent, when applied to some special case is not a familiar one. Yet it is easy to construct such rules once the possibility is raised.

[4] One might argue that this paradox is not, strictly speaking, a semantic paradox, on the grounds that 'eligible to join' is not a semantic relation. *But the expression 'semantic paradox' has come to denote any paradox of the sort Russell called 'vicious circle paradox' that is not purely logical or mathematical in nature, and it is for this reason I call the Sec Lib a semantic paradox.*

(593-595 in 'The Semantic Paradoxes: A Diagnostic Investigation'. My first and last emphases.)

This extended quote reveals what I regard as a thoroughgoing confusion on Chihara's part. His argument for the rejection of (1) strikes one as perfectly reasonable, indeed as showing that we do *not* have here a genuine semantic paradox. His psychologistic explanation subsequently is quite irrelevant to the problem of giving *objective, logical* criteria for paradox, which is of course what I am attempting to do by using the techniques of proof theory. It is the footnote that provides some explanation of Chihara's mistake in regarding Sec Lib as a paradox. Where, in the case of Sec Lib, is the *vicious circle* that allegedly turns it into a genuine semantic paradox, absence of semantical predicates notwithstanding? I see none (and indeed the proof theoretic analysis given below reveals none). Is it because the reasoning to absurdity ends with a little "iff not" punch line? If so, the whole diagnosis is not at all convincing. The proof of Cantor's Theorem ends similarly, but noone regards it as viciously circular. This is an example of *inchoate* semantical intuitions concerning relatively *trivial* data being blown up into misguided diagnostic theory — rather like the student of logic who, upon learning that one cannot infer P from — —P in intuitionistic logic, immediately concludes that the converse inference is not intuitionistically valid either.

Nor does any vicious circularity arise from the fact that Sec Lib's rules — rules that are constitutive of Sec Lib itself — impredicatively generalise over the class of all clubs, taken of course to contain Sec Lib itself. Intuitively, the 'paradox' reveals nothing more troublesome than that the eligibility rules of Sec Lib are incoherent if taken seriously as *general* rules. (Must one repeat the truth that just one counterexample, however remote in space and time, however infrequent in daily life, serves to overthrow a universal generalisation? Chihara's psychologism on this score alters this not one jot.)

If my proposed criterion for paradoxicality is correct, then the following proof of the Sec Lib 'paradox' shows it not to be a genuine semantical paradox. That the informal reasoning tails off with an "iff not" is not enough to wag the dog of paradox.

We have the eligibility rule that any person t is eligible to join c (Sec Lib) iff t is secretary of some club that t is not eligible to join. This may be expressed by the two deductive rules

$$\dfrac{Etc \qquad \begin{array}{c} \overset{(i)}{\overline{Sta}} \quad \overline{-Eta}^{\,(i)} \\ \vdots \\ \psi \end{array}}{\psi}\,(i) \qquad\qquad \dfrac{Stu \qquad -Etu}{Etc}$$

where a does not occur in ψ, or any assumption, other than Sta and —Eta, on which the upper occurrence of ψ depends

Consider now the following:

SEE DIAGRAM 11

Our final proof is in normal form. Thus no more than the bare inconsistency of the Sec Lib scenario has been established. There is not the faintest hint of genuine semantical paradox. The eligibility rules are simply ill-conceived, even if this is brought out only by the special case of Ms Fineline. Granted, no Aristotelian continuants have inconstistent properties, while on the other hand social individuals such as clubs and other institutions might have inconsistent constitutions; but *this* fact of life cannot elevate Sec Lib to the level of a semantical paradox.

5. *Is paradoxicality an absolute or a relative notion? And what is it a property of?*

Kripke has recently stressed that

> many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavourable, to exhibit paradoxical features.
> ('Outline of a Theory of Truth'. p 691)

This can be most simply illustrated by what I call the Wedge Liar, and is known also as the postcard paradox. Consider utterances of

(1) The assertion α by A is false
(2) The assertion β by B is false

where α and β are empirical descriptions so formed that, given the way the world is — that is, how people have made assertions in and about it — the assertion α by A is an assertion of (2) and the assertion β by B is an assertion of (1). The schematic form is thus

$$\beta \dashv\vert\vert\vdash -T\alpha$$
$$\alpha \dashv\vert\vert\vdash T\beta$$

where $\dashv\vert\vert\vdash$ must now be understood as relative to the way, referentially, things happen to work out. The reader may be curious to see how the present example fares on my test. Bearing in mind the truth schema as before, we have:

SEE DIAGRAM 12

If the reader writes out the final proof in full, and tries to normalise it by shrinking branches of the form $\gamma\ldots\gamma$ to γ, and applying the reduction procedure for negation, he will find that the resulting reduction sequence loops.

In going on to discuss the notion of *grounded* truth and falsity, Kripke remarks (p 694)

> whether a sentence is grounded is not in general an intrinsic (syntactic or semantic) property of a sentence, but usually depends on the empirical facts.

Paradoxicality, likewise, is relative to the empirical facts. For Kripke defines a sentence as paradoxical if it has no truth value at any fixed point. (I assume that the reader is familiar with this notion from Kripke's paper.) Now the interpreted languages in question (including the fixed points) all extend $L_O =_{df} L(\emptyset, \emptyset)$, the interpreted language in which the extension and anti-extension of the truth predicate are both empty. In L_O, however, all the 'empirical' predicates are *fully interpreted.* L_O may therefore be thought of as a model ("the empirical facts"). Kripke's definition therefore states that a sentence is paradoxical *relative to this model* if it has no truth value at any fixed point extending it.

A natural extension of this definition suggests itself. Let us call a *set* of sentences (such as the pair of sentences for the Wedge Liar) paradoxical just in case there is no fixed point (extending L_O) at which every member of the set receives a truth value. Paradoxicality, on this semantic account, is thus a property of sets of sentences relative to models.

Let Θ be the set of sentences of L_O that are true and free of the truth predicate. A question that immediately arises is whether paradoxicality is so sensitive to the exact nature of the model in question that Θ (especially non-categorical Θ) cannot do duty for it. Can paradox turn on empirical features so fine that elementarily equivalent but non-isomorphic models would yield different results for the paradoxicality or otherwise of certain sets? Another question that arises is whether the semantical notion of paradoxicality is compact. Does every paradoxical set (relative to a given model) include a paradoxical *finite* subset (relative to that model)? These are questions that arise by extending Kripke's initial suggestions on the semantic side by analogy with the more familiar notions of consistency (or satisfiability) and consequence in semantically open languages. But by far the most interesting question is whether we can *axiomatise* the notion of paradoxicality. To this we now turn.

6. *A completeness conjecture on paradoxicality*

We shall say that a set of sentences is paradoxical relative to a class of models if it is so relative to each member thereof, and is *purely* or *intrinsically* paradoxical if it is paradoxical relative to every model.

How now does one connect these ideas with those concerning proofs? Notice how in our example of the Wedge Liar the empirical facts (the model) made β refer to (thus 'interdeducible' with) $\neg$Ta, and made a refer to (thus 'interdeducible' with) $T\beta$. These *id est* inferences are an important part of $\Theta(M)$, which I propose to treat both sententially and inferentially. When, therefore, I speak of a proof of a conclusion from $\Theta(M)$, I have in mind a proof from assumptions that are truths in (every member of) M, and by means of rules for the logical operators and the truth predicate, as well as the *id est* rules of inference — really, one could say, rules of *reference* — that are legitimated by (every member of) M. The completeness conjecture is then that

> A set of sentences is paradoxical relative to M
>
> iff
>
> there is some proof of Λ from $\Theta(M)$, involving those sentences in *id est* inferences, that has a looping reduction sequence.

Compactness of paradoxicality would follow as a corollary. So would the view that paradox is really only theory-sensitive and not model-sensitive.

Clearly some result such as this is worth having, given the ideal and infinitistic nature of Kripke's semantical definition of paradox. Indeed, when Kripke shows how the usual examples, such as the Liar, come out paradoxical on his definition, he is doing no more than giving a metalinguistic version of the reasoning that, as we have seen, cannot assume the form of a normal proof. Thus Prawitz's objections to the "black box" aspect of the classical definition of consequence apply here, with suitable modifications. (See his paper 'On the Idea of a General Proof Theory'.)

7. *Reflections on normalisability and meaning theory*

In the proof based meaning theory canvassed by Prawitz, normalisability of proofs is an important adequacy condition on the account of the meanings of those operators and constructions that have their own introduction rules. An operator with introduction and elimination rules such as to block normalisability in general is liable to be semantically suspect on this account. Moreover, since we have in the non-normalisable proofs considered above an easily identifiable suspect — the rules for T — does this not just underline, from the point of view under discussion, the incoherence of truth predication within the object language? Is not the existence of these looping reduction sequences further evidence that semantic closure is intolerable, a semantics for closed languages unattainable?

I think not, and my reason is as follows. The importance of normalisability from an intuitionist point of view is its role as a soundness proof for the logical system. Normalisability guarantees the existence of an effective method, given any proof of a conclusion from (possibly complex) premisses, to transform canonical proofs of the latter (relative to any base) into a canonical proof of the former (relative to that base).

Now in the case of a non-normalisable paradox proof relative to M what we have essentially is a welcome *block* to any proof of Λ relative to M. If the proof were normal, and if normalisability held generally, then any canonical proofs for its undischarged assumptions drawn from Θ(M) (relative to any base) would be transformable to a canonical proof of Λ relative to that base. (I am of course assuming that if the proof

$$\begin{array}{c}\underbrace{\psi_1, \ldots, \psi_n}\\ \Pi \\ \psi\end{array}$$

can't be normalised, then nor can any $\begin{array}{c}\underbrace{\begin{array}{ccc}\Pi_1 & & \Pi_n\\ \psi_1, & \ldots, & \psi_n\end{array}}\\ \Pi\\ \psi\end{array}$, Π_i canonical.)

The impossibility of ascribing truth values to the sentences shown by a proof to be paradoxical relative to M saves M itself from inconsistency.

In a semantically closed language, therefore, we should not want a normalisation theorem. This is moreover eminently compatible with trusting as genuine guarantors of consequence only those proofs that are normal, or that can be brought into normal form.

The general loss of normalisability, confined as it is according to our conjecture above to just the paradoxical part of the semantically closed language, is a small price to pay for the protection it gives against paradox itself. Logic plays its role as an instrument of knowledge only insofar as it keeps proofs in sharp focus, through the lens of normality. Normalisability, in the context of semantically closed languages, is not to be pressed as a general pre-condition for the very possibility of talking sense; rather, normality of proof is to be pressed as a general pre-condition for the very possibility of telling the truth.

A similar theme could be developed in the light of the proof theoretical results about entailment referred to above. What I have called Proofs (ie proofs in normal form, with no applications of the absurdity rule) can be accumulated in transitive fashion to produce overall a Proof *either* of the desired conclusion *or* of absurdity. To discover which is indeed the case — and there is an enormous epistemic difference between the alternatives — we have to perform certain operations on the overall proof, first normalising it and then winnowing out applications of the absurdity rule that might materialise during the normalisation process. It is the finished Proof that then gives the reliable result. And so it is with paradox too.

I undertook at the beginning to confine my attention as much as possible to intuitionistic proofs. It is quite clear that nothing we have seen in the paradoxes turns remotely on strictly classical features of the underlying logic. It appears to me an open question whether every paradoxical set of sentences (relative to a model) can be shown to be paradoxical by means of an *intuitionistic* proof with a looping reduction sequence.

Dana Scott has suggested an apparent counterexample to this conjecture. He points out that in a language treating functions as on a par with their inputs and outputs, the sentence $\sim 0 = 1$ and the choice schema $\forall x \exists y\, \varphi(x, y) \supset \exists f\, \forall x\, \varphi(x, fx)$ form an intuitionistically consistent but classically *'paradoxical'* set. But the air of paradox in the classical case is only apparent, suggested perhaps by the diagonal nature of the reduction proof. This proof turns out to be in normal form, and therefore on my account reveals straightforward inconsistency rather than paradox.

SEE DIAGRAM 13

Notes

I understand from Jaakko Hintikka that he is developing a game theoretical approach to the paradoxes, and finds that their distinctive feature is that the associated game enters an infinite sequence of moves, never resulting in a "win-lose" end position. This obviously bears some relation to the proof theoretic ideas presented here, and it remains to investigate the connection more closely. Also, Hans Herzberger has modified Kripke's method of truth value assignments to semantically closed languages, so that a sentence's truth value can change back and forth up the hierarchy. He proves that there comes a point beyond which the pattern of change is periodic. Again, this might tie in with the properties of proofs discussed above. Priest maintains that the reasoning in the definability paradoxes, such as Berry's, does not display the pattern I am suggesting as characteristic of paradoxes; the structure of the proofs concerned should therefore be investigated more closely.

DIAGRAM 1

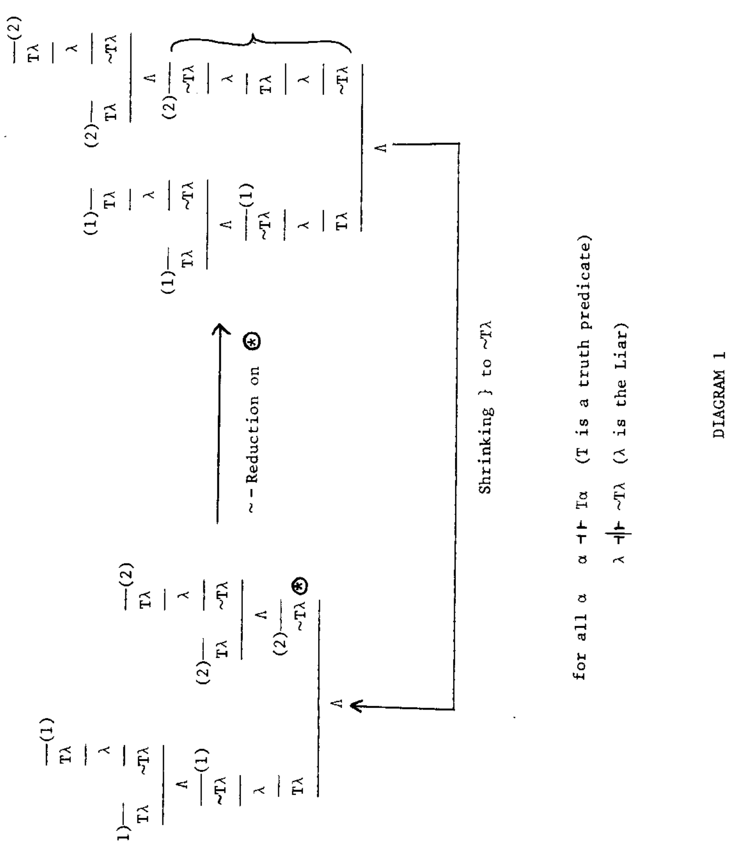


DIAGRAM 2

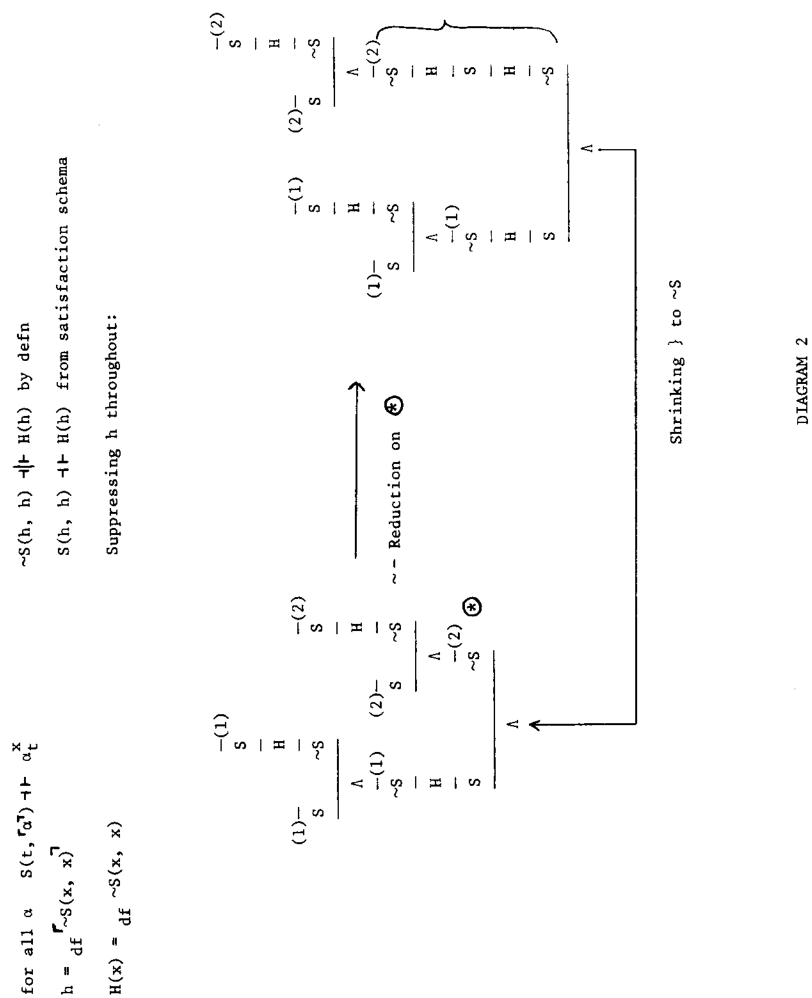


DIAGRAM 3

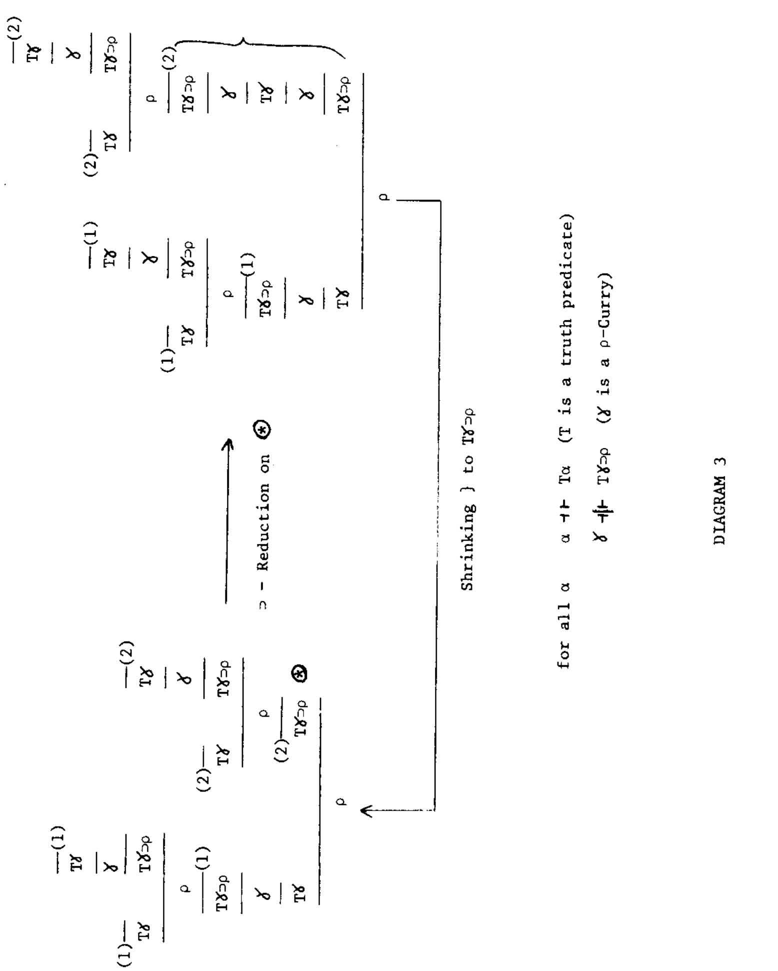


DIAGRAM 4

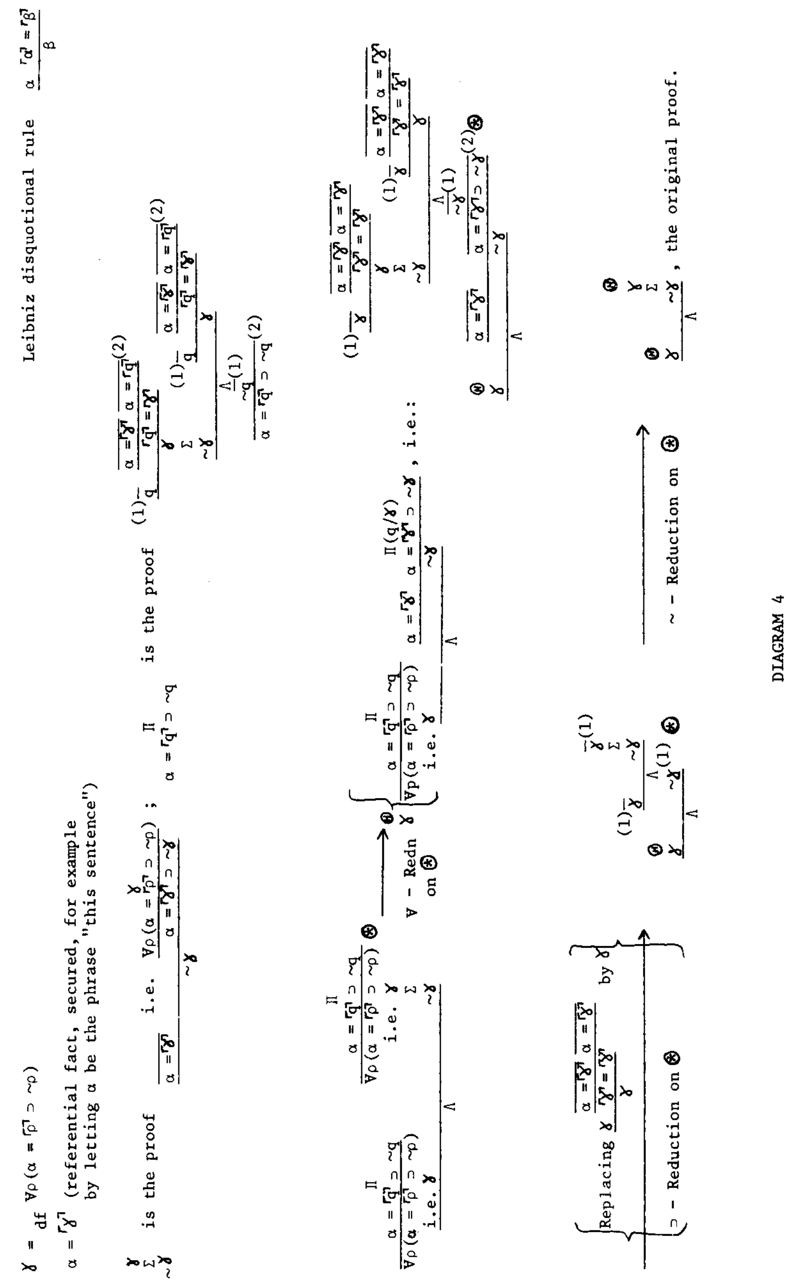


DIAGRAM 5

$$\Pi\varepsilon \quad \frac{\psi^x_t}{t\varepsilon\{x|\psi\}} \qquad \varepsilon\Pi \quad \frac{t\varepsilon\{x|\psi\}}{\psi^x_t}$$

$$\rho =_{df} \{x|\sim x\varepsilon x\}$$

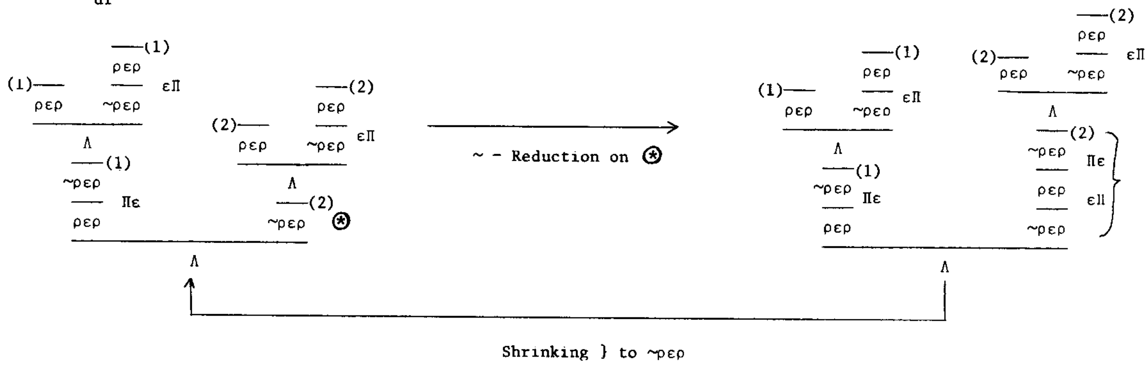


DIAGRAM 5

DIAGRAM 6

$$\Pi\varepsilon' \quad \frac{\psi^x_t \quad \exists!t \quad \exists!\{x|\psi\}}{t\varepsilon\{x|\psi\}} \qquad \varepsilon\Pi \quad \frac{t\varepsilon\{x|\psi\}}{\psi^x_t}$$

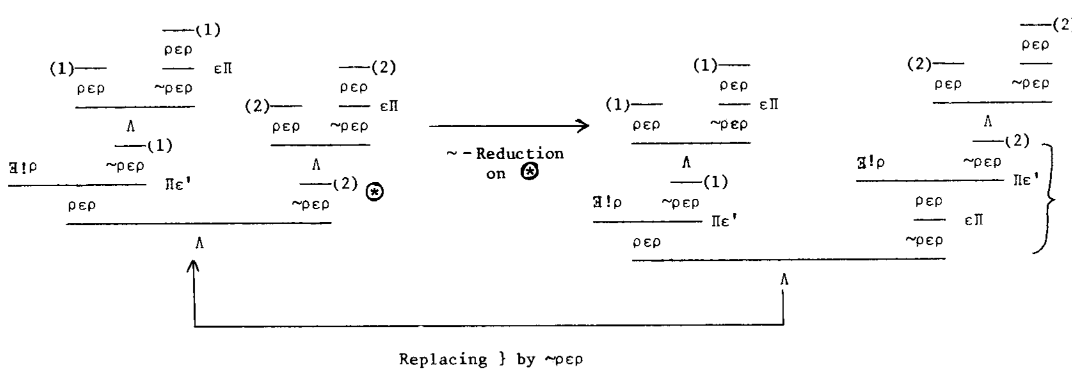


DIAGRAM 6

DIAGRAM 7

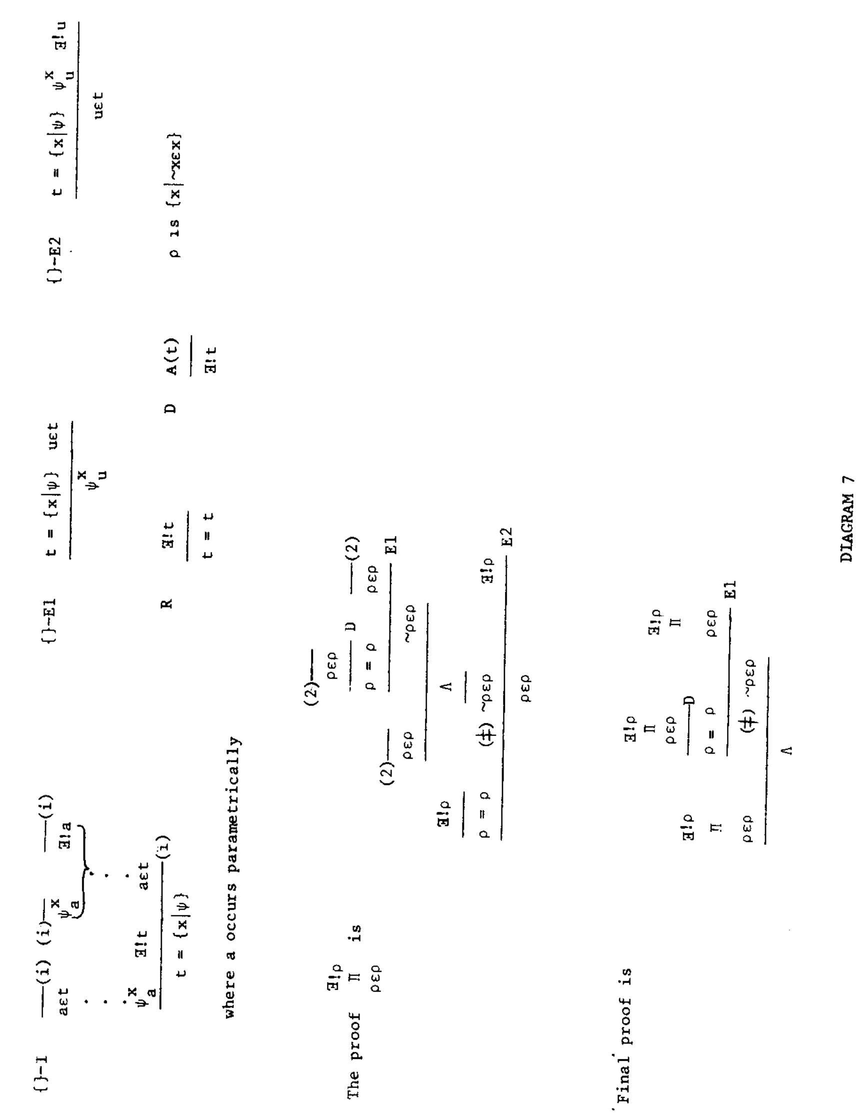
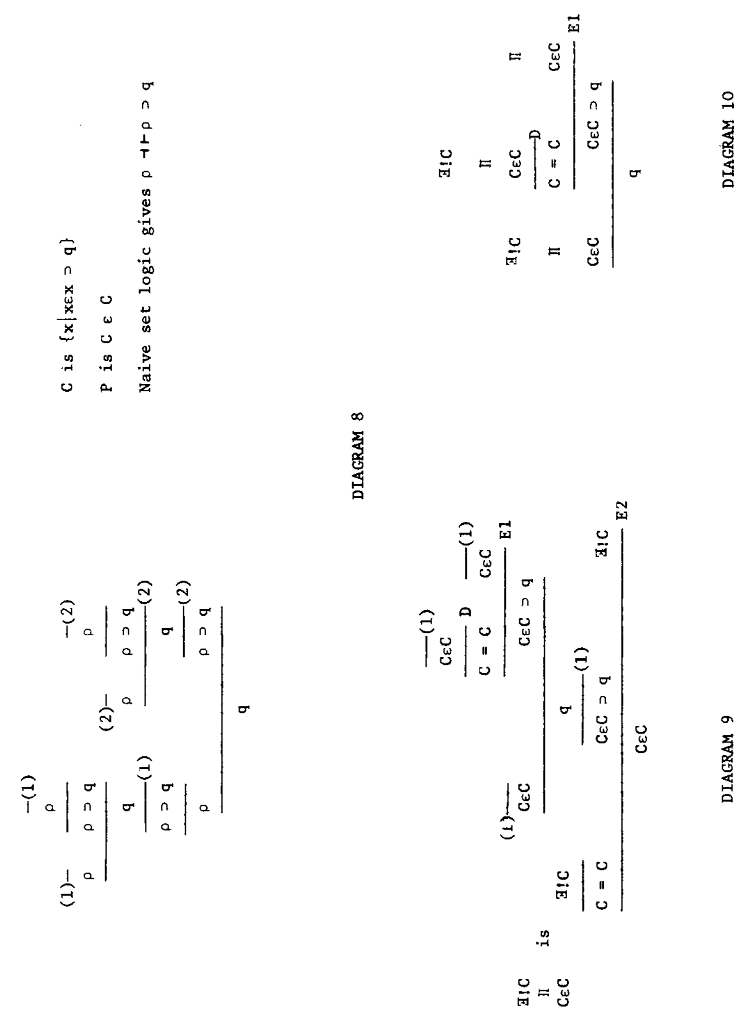
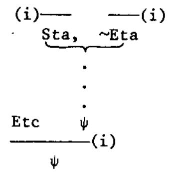
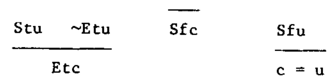


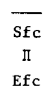
DIAGRAM 8, 9, 10



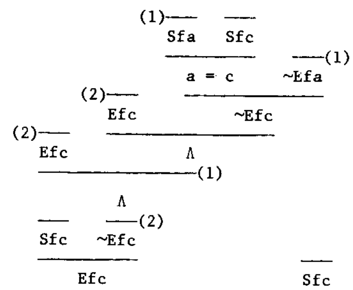




where a is parametrical



is the proof



The final proof (in normal form!) is

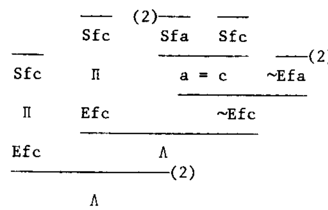


DIAGRAM 11

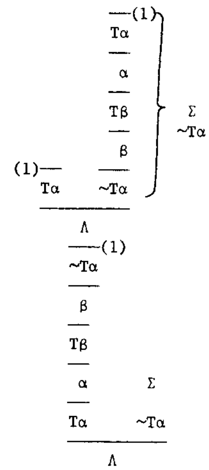
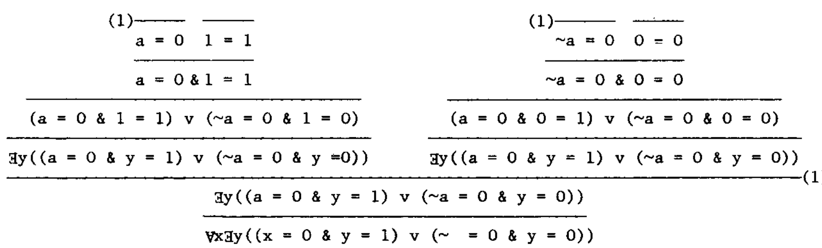


DIAGRAM 12

φ(x, y) : (x = 0 & y = 1) v (~x = 0 & y = 0)

Proof that ∀x∃yφ(x, y)

Classical



Proof that ∃f∀xφ(x, fx) ⊢ Λ

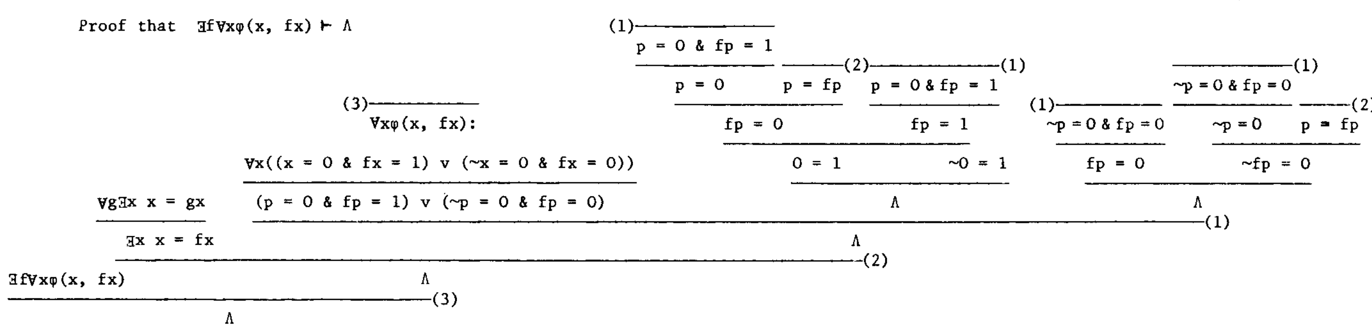


DIAGRAM 13

REFERENCES

R.L. Martin and P.W. Woodruff. 'On representing "true in L" in L', in ed A. Kasher, *Language in Focus* (Reidel 1976) pp. 113-117.

S. Kripke. 'Outline of a theory of truth', *Journal of Philosophy* 62 (1975) 690-716.

C. Chihara. 'The semantic paradoxes — a diagnostic investigation', *Philosophical Review* 88 (1979) 590-618.

N. Tennant. *Natural Logic* (Edinburgh, 1978); 'A proof-theoretic approach to entailment', *Journal of Philosophical Logic* 9 (1980) 185-209.

D. Prawitz. *Natural Deduction — A proof theoretical study* (Stockholm, 1965).

D. Prawitz. 'On the idea of a general proof theory', *Synthese* 27 (1974) 63-77.

R.K. Meyer, R. Routley and J.M. Dunn. 'Curry's Paradox', *Analysis* 40 (1980) 124-128.

G. Takeuti and W.M. Zaring. *Introduction to Axiomatic Set Theory*, Springer (1971).

Vol. 36
Fasc. 4
1982

dialectica

Publié avec l'aide de la Société Suisse des Sciences Humaines (Académie Suisse des Sciences Humaines)
Publiziert mit Unterstützung der Schweizerischen Geisteswissenschaftlichen Gesellschaft (Schweizerische Akademie der Geisteswissenschaften)

International review
of philosophy
of knowledge

Revue internationale
de philosophie
de la connaissance

Internationale
Zeitschrift
für Philosophie
der Erkenntnis