THE REVIEW OF SYMBOLIC LOGIC
Volume 18, Number 4, December 2025

CONSTRUCTIVE ARITHMETICAL IMPOSSIBILITIES AND THEIR RELATION TO PARADOXES

NEIL TENNANT
The Ohio State University

Abstract. This study focuses on certain combinations of rules or conditions involving a would-be 'provability' or 'truth' predicate that would render a system of arithmetic containing them either *straightforwardly inconsistent* (if those predicates were assumed to be definable) or *logico-semantically paradoxical* (if those predicates were taken as primitive and governed by the rules in question). These two negative properties are not to be conflated; we conjecture, however, that they are complementary. Logico-semantic paradoxicality, we contend, admits of proof-theoretic analysis: the 'disproofs' involved do not reveal straightforward inconsistency. This is because, unlike the disproofs involved in establishing straightforward inconsistencies, these paradox-revealing 'disproofs' cannot be brought into normal form.

The border between metamathematical proofs of certain (constructive) impossibility results and the non-normalizable (and always constructive) disproofs engendered by semantic paradoxicality is not fully understood. The respective strategies of reasoning on each side—genuine proofs of inconsistency *versus* whatever kind of 'disproof' uncovers semantic paradoxicality—seem somehow similar. They seem to involve the same 'lines of reasoning'. But there is an important and principled difference between them.

This difference will be emphasized throughout our discussion of certain arithmetical impossibility results, and closely related semantic paradoxes. The proof-theoretic criterion for paradoxicality is that in the case of paradoxes (as opposed to genuine inconsistencies) the apparent 'disproofs' that use the rules stipulated for the primitive predicates in question cannot be brought into normal form. In proof-theoretic terminology: their reduction sequences do not terminate. This means that CUT fails for languages generating paradox. But CUT holds for the language of arithmetic. It follows that the paradox-generating primitive predicates of a semantically closed language cannot be defined in arithmetical terms. For, if they could be, then they could be replaced by their definitions within the paradoxical disproofs, and the resulting disproofs *would* be normalizable.

§1. Introduction. The border between metamathematical proofs of certain (constructive) impossibility results and semantic paradoxicality is not fully understood. The respective strategies of reasoning on each side are genuine proofs of inconsistency *versus* whatever kind of 'disproof' uncovers semantic paradoxicality. They seem somehow similar. They seem to involve the same 'lines of reasoning'. But there is an important and principled difference between them.

Received: July 1, 2024.
2020 *Mathematics Subject Classification*: Primary 03A05, 03B20, Secondary 03F99.
Key words and Phrases: arithmetical impossibilities, paradoxes, constructive logic.

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Association for Symbolic Logic. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

doi:10.1017/S175502032510097X

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

This difference will be emphasized throughout our discussion of certain arithmetical impossibility results, and closely related semantic paradoxes (actual or conjectural). The proof-theoretic criterion for paradoxicality is that in the case of paradoxes (as opposed to genuine inconsistencies) the apparent 'disproofs' establishing absurdity cannot be brought into normal form. Proofs and disproofs that are in normal form have no sentence occurrence standing as the conclusion of an application of the introduction rule for its dominant logical operator and as the major premise of an application of that operator's corresponding elimination rule. Each operator has a well-known reduction procedure that ensures the removal of any such 'local peak' of unnecessary complexity.[1] In proof-theoretic terminology: their reduction sequences do not terminate.

This entails that CUT fails for languages generating paradox. But CUT holds for both constructive and classical logic for the language of arithmetic. It follows that the paradox-generating primitive predicates of a semantically closed language cannot be defined in arithmetical terms. For, if those predicates (the *definienda*) *could* be so defined, then they could be replaced by their definitions (the arithmetical *definientia*) within the paradoxical disproofs employing those predicates, and the resulting disproofs, which would be employing purely arithmetical expressions in their sentences, *would* be normalizable.

§2. Paradoxicality vs. genuine inconsistency. The meta-rule of CUT:

$$\text{CUT} \qquad \frac{\Delta \vdash_{\mathcal{S}} \varphi \qquad \Gamma, \varphi \vdash_{\mathcal{S}} \psi}{\Delta, \Gamma \vdash_{\mathcal{S}} \psi}$$

expresses the unrestricted transitivity of deduction within the system $\mathcal{S}$ in question. The underlying logic for the language of first-order arithmetic of course enjoys such transitivity; and one can freely invoke this meta-rule within any axiomatized system of first-order arithmetic. As we shall see, we can invoke the constant applicability of CUT within such systems $\mathcal{S}$ in order to show that $\mathcal{S}$ cannot, on pain of contradiction, contain any 'provability' or 'truth' predicates (of code numbers of sentences) meeting certain conditions.

There is an interesting variety of combinations of conditions that will precipitate the inconsistency of $\mathcal{S}$. *Nota bene*: these will be *genuine* inconsistencies, established by *dis*proofs within $\mathcal{S}$ (i.e., $\mathcal{S}$-proofs of $\perp$) that are *in normal form*. And what is especially noteworthy is that the disproofs in question are constructive. That is, they can be formalized in Intuitionistic Logic. Indeed, they can be formalized in Core Logic, since the reasoning involved in them is relevant (as is the case with all informally rigorous mathematical reasoning, be it constructive or strictly classical). Collectively, the results establishing the constructive inconsistency of contemplated combinations of conditions on 'provability' predicates mark out a frontier for systems that can represent their own syntax—a frontier beyond which, should one make so bold as to allow any new *primitive* predicates to behave in these forbidden ways, one would be confronted (so we conjecture) with the dragons of paradox.

[1] This criterion was first advanced in [7] (see also [8–10, 12]).

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

The (constructive) *arithmetical impossibility* results tell us

> You cannot *define predicates or operators* within your existing language for arithmetic that conform to such-and-such deducibility conditions; for if you *had* such definitions they would precipitate, constructively, *inconsistency*. And those proofs of inconsistency would be *normalizable*—i.e., able to be brought into normal form by applying reduction procedures for the logical operators involved, *á la* Gentzen and Prawitz.

The corresponding logico-semantic *paradoxicality* results would tell us

> You cannot *extend* your language for arithmetic with primitive predicates or operators and stipulate that they are to obey rules of inference codifying those deducibility conditions; for by the same 'lines of argument' as in the case of the attempted definitions mentioned in the previous paragraph, use of those rules will lead, constructively, to *logico-semantic paradox* within your extended language. That is, they will allow there to be disproofs (proofs of $\perp$, i.e., absurdity) that *cannot* be brought into normal form. The reduction sequences generated by applying the reduction procedures for the logical operators involved *will not terminate*.

Our emphasis on constructivity in this regard is doctrinal. Our view is that paradoxes inflict their intellectual pain constructively. They reveal problems with our conceptual apparatus which lie deep and do not require allegiance to Bivalence to be winkled out (*via* any of Bivalence's attendant strictly classical inference rules known as the Law of Excluded Middle, the rule of Double Negation, the rule of Classical Reductio, or the rule of Classical Dilemma). Likewise, the metamathematical and arithmetical reasoning for the results analyzed in this study can all be arrived at *constructively*; and it is the main purpose of this study to demonstrate this, at the level of formal rigor required. The corresponding reasoning in the paradoxical setting (where the problematic predicates are taken as primitives) will then, not surprisingly, turn out to be constructive also. The reduction sequences for the formal disproofs will not terminate—or so we conjecture. For lack of space we prescind here from the latter details, but hold them in morally certain prospect. We believe that what happens with the Liar (see below) *generalizes* to the other logico-semantic paradoxes.

There is a *topical divide* here, which is already illustrated clearly in the case of the arithmetical indefinability of a truth predicate for arithmetic *versus* the Liar paradox, which arises from taking such a predicate as a *primitive*. On one side of the topical divide are the constructive *arithmetical impossibilities*; on the other side of the divide are the complementary logico-semantic paradoxes that would arise (so we conjecture) by taking the predicates involved not as arithmetical *definienda*, but rather as primitive operators in their own right, governed by the conditions stated for the impossibility results. There is, we contend, a deep correspondence between the two side of the topical divide.

In our discussion thus far we have spoken of arithmetical impossibility results and then segued into mention of the associated or 'somehow related' paradoxicality that arises when one treats the problematic predicates as primitives and therefore does

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

not assume them to be explicitly definable. Historically, of course, matters presented themselves in the reverse order. The ancient philosopher Eubulides (*circa* 405–300 BCE) initiated the lesson that we cannot add primitive predicates (such as 'is true') to a language containing expressions denoting sentences and stipulate that those primitive predicates are to obey rules such as the obvious-looking 'φ; *ergo*, $\ulcorner\varphi\urcorner$ is true' and its converse '$\ulcorner\varphi\urcorner$ is true; *ergo*, φ'.[2] This is because of the notorious Liar sentence λ, which says '$\ulcorner\lambda\urcorner$ is not true'. If one tries to regiment the deductive reasoning in the resulting Liar paradox as a disproof, it turns out that the disproof is not in normal form, and is non-normalizable.[3] The ancients of course could not conceptualize matters this way, because they lacked the resources of modern Gentzenian proof theory. The proof-theoretic observation here is an explication of the intuited vicious circularity induced by the combination of self-reference with over-hastily adopted rules of inference governing the truth predicate.

Well over two millenia later, this lesson was reprised (or *applied*) rather differently, in the context of metamathematical studies of the first-order language of arithmetic, in which consistent and *sufficiently strong* arithmetical theories $\mathcal{S}$ afforded representation of linguistic expressions via numerical coding. This enabled those theories to furnish linguistic representation of decidable relations among such expressions, via representing formulae involving numerals for the code numbers of those expressions. *This is the basic contemporary requirement for a consistent arithmetical theory to be called 'sufficiently strong'*. Somewhat ironically, the rather weak (in the arithmetical sense) theories R and Q—due to Raphael Robinson—turn out to be sufficiently strong in this 'logical' sense. The lesson was that one would not, *on pain of inconsistency*, be able to *define* (explicitly) that same predicate ('is true') is any such system $\mathcal{S}$. The deductive reasoning involved is straightforwardly mathematical, and can be regimented as a disproof in normal form. What we would informally call the *line of argument*, however, is that of the ancients. In a nutshell: Tarski employed the same line of reasoning as Eubulides. But what they respectively did was subtly different. Eubulides revealed the *logico-semantic paradoxicality* of having a truth predicate as a primitive predicate, governed by intuitively appealing and seemingly obvious rules, in a language permitting reference to its own sentences. Tarski, by contrast, revealed the *impossibility of defining* a truth predicate in arithmetical terms, in a language for arithmetic that afforded numeralwise representation of its own expressions. No would-be definiens in the language of arithmetic would work. The assumption that any would-be definiens *would* work provably implies *straightforward inconsistency*, not logico-semantic paradoxicality.

The later results in this initially Tarskian vein, which we shall be studying here, and which are due to Löb, Montague, and McGee are (like Tarski's), constructive (or, if not constructive upon first presentation, *constructivizable*). But these classically-minded metamathematicians did not have any axe to grind over constructivity. They produced their deep impossibility results (about the indefinability of arithmetical predicates satisfying certain sets of plausible-looking conditions—the set in question depending on the author) with no explicitly expressed concern at all to have demonstrated them constructively. But when matters are 'deeply constructive', it will often be the case that

[2] Eubulides, of course, did not formulate the problem in such formal terms; we are taking the liberty of so doing, for him.

[3] We demonstrated this in [7].

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

habitually non-constructive mathematical reasoners will happen, in their initial foray, upon what turn out to be constructive proofs.

We wish to take the necessary pains here to bring out the kernel of constructivity in all of these impressive results. We intend thereby to lend further plausibility to the organic connection (the recurring theme to which we wish to draw attention) between the impossibility of would-be arithmetically defined predicates, on the one hand, and the logico-semantics paradoxicality, on the other hand, of venturing to have primitive 'surrogates' of them, naively assumed to be governed by the same plausible-looking conditions. We are obliged to postpone to a sequel to this study the painstaking demonstration of paradoxicality in the 'primitive', as opposed to the 'arithmetical', setting, for the results of Löb, Montague, and McGee. For these the historical order that we have commented on in connection with the Liar would have to be reversed. There are no known, let alone well-known, logico-semantic paradoxes that would correspond with the impossibility results of Löb, Montague, and McGee in the way we are contending. We would have to take their results, fully formalized as disproofs, and 're-do' them as disproofs (paradoxical ones) in a language that purports to treat the problematic predicate as a primitive governed by rules of inference codifying the sets of conditions that these authors, respectively, listed; and then we would have to show that the disproofs are not normalizable.

§3. Our chosen case studies. We shall investigate here some important impossibility results about first-order systems of arithmetic with 'provability' or 'truth' predicates (defined or postulated). We rigorously formalize the proofs of these impossibility results in the natural deduction system for Intuitionistic Logic. They are proofs of straightforward, non-paradoxical, *inconsistency*. As already intimated, we leave for a sequel the parallel treatment (or analysis) of the same cases when their *paradoxicality* is at issue. That treatment would involve the regimentation of disproofs, and the demonstration that their reduction sequences do not terminate, when each of these results is re-cast in the guise of 'adopting a primitive predicate' governed by the rules of inference that codify the problematic conditions on that predicate.

The Liar Paradox is by far the best known logico-semantic paradox to which our recommended proof-theoretic 'paradoxicality analysis' has been applied. And there are a host of other well known logico-semantic paradoxes that likewise, respectively, admit of such paradoxicality analyses.[4] We are not, however, looking for any 'arithmetical inconsistency' results corresponding to these. Our interest here is in the reverse correspondence: will the arithmetical inconsistency results of Löb, Montague, and McGee engender logico-semantic paradoxes that will admit of paradoxicality analyses of the kind we have described?

We are not aiming for comprehensiveness; there could well be other results of this kind that the reader could use to test the general theme that we are trying to illustrate with our chosen examples. Even so, we do not think we can fairly be accused of 'cherry picking'; for these are the most striking results of this general kind that came to the author's mind when the 'general picture' came to him. Moreover, the results are all

[4] Among these are the Grelling Paradox, the Postcard Paradox, the Curry Paradox, Yablo's Paradox, Prior's Paradox, and the Revenge Paradox.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

deep and impressive and important. So there is an element of desert—of homage to the literature—if they are indeed cherries that have been picked.

The results we shall focus on are:

1. Tarski's theorem about the indefinability of truth in arithmetic (see §5);
2. Löb's theorem about limitations on the (defined) provability predicate in arithmetical systems (see §6);
3. Montague's theorem about how the necessity operator cannot be treated as a predicate of sentences (see §7);
4. McGee's version of the Liar, using weaker conditions on the would-be truth predicate (see §8);
5. McGee's theorem about the ω-inconsistency of certain natural-looking (postulated) conditions on a provability predicate in arithmetical systems (see §9).

A summary of these five results is provided in §10. The reader might find it helpful to look ahead to §10 for an overview, and also to consult it from time to time as our study progresses.

(1) is the simplest of these impossibility results and we have already discussed it in an introductory fashion in §2. We shall revisit it in §5 and—for McGee's strengthened version (4) of the result—in §8. (1) keeps company with (2)–(5) and has inspired the present more detailed study of the latter four. The well-known Liar Paradox of Eubulides became, in Tarski's hands (as we have already mentioned in §2), the result that arithmetic cannot contain its own (defined) truth-predicate.[5] With (2)–(5), the scholarly segue will have to be in the reverse order.

Once we have collected these inconsistency results, and established them constructively, we shall have learned something important about the anatomy of the aforementioned dragons. The lesson will be: for languages allowing the forbidden behavior on the part of certain of its predicates, *the meta-rule of* CUT *fails*. It has been argued elsewhere that that very failure is the mark of paradox (as opposed to genuine inconsistency). It has been argued also that the deductive reasoning within the paradoxes that are involved is wholly constructive. That is, paradox cannot be blamed on any allegedly noxious contribution of strictly classical reasoning. The sources of paradox afflict the constructivist just as seriously as they do the classicist.

The remaining aim in this investigation, given limitations of space, is to *constructivize* the deductive reasoning in the important 'arithmetical impossibility' results (2)–(5) that are described and established in more detail below. Such results show the impossibility of having *explicit definitions* of provability- or truth-predicates satisfying certain collections of conditions. The proofs of these impossibilities are—to repeat—constructive, and involve 'CUTs galore'. The CUTs are permitted because we have CUT-Elimination[6] for the first-order language of arithmetic, in which, one is assuming, those explicit definitions are formulated.

[5] The proof of this latter result has already been formalized (in Core Logic) in [13], whose title is 'Core Tarski and Core McGee'. The McGee result that [13] discusses (about Tarskian biconditionals) is a different one from the just-mentioned McGee results (4) and (5) to be discussed here.

[6] Indeed, CUT-*Admissibility* would suffice.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Shift now, then, to imagining that instead of trying to *define* the provability-like or truth-like predicates so as to satisfy the various collections of conditions, one simply adopts such (new) predicates as *primitive*, and as governed by rules of inference that give full and faithful expression to those collections of conditions. One will then—we contend—be confronted by (genuine) *paradox*, in the form of stretches of deductive reasoning that appear to establish absurdity from no sentential premises, but only by the action of the rules adopted for the new primitive predicate in question. This entails that CUT must fail for the system of reasoning that contains the rules in question—on pain of having the (augmented) system of rules itself be inconsistent. And that failure of CUT manifests itself as the failure of attempts to *normalize*, i.e., *eliminate the* CUT*s from*, the formal proofs regimenting the deductive reasoning involved.

The reader might well wonder why the technical work here devoted to furnishing explicitly constructive proofs is at all necessary, in light of the general result of [2] that Heyting Arithmetic and (classical) Peano Arithmetic prove the same Π^0_2-sentences. Since the arithmetical impossibility results (construed as theorems of the form $\neg\varphi$) are Π^0_2, why go to all this bother? The answer is that we wish to avoid using any hammer to crack these few walnuts because we want the edible flesh to be in pristine condition, as manifestly constructive proofs, when it comes to re-framing them as the proofs involved in the manifestation of paradoxicality, once the predicate at issue is taken to be a primitive, rather than being assumed (for *reductio*) to be explicitly definable in the language. *That* will involve translation of the explicit formal proofs we will have found; we will not be able to rely upon the mere assurance that they exist.

The deductive, rule-governed reasoning with a paradox does not establish a genuine inconsistency; the 'disproof' involved cannot be brought into normal form. This point can be illustrated most simply with the Liar. As already mentioned we have taken the trouble to provide that more detailed illustration in our earliest work on paradox. And the Liar is treated in the same fashion in the later works cited in footnote 1 on p. 2. Similar demonstrations of the non-normalizability of the 'disproofs' involved in the 'primitive predicate' versions of the other impossibility results (2)–(5) will be left to the intrigued reader as tantalizing and non-trivial exercises (and perforce by the present author to a sequel).

§4. Systems interpreting Robinson arithmetic Q**.** We turn our attention now to formal first-order systems for arithmetic, since that is the setting for all the constructive impossibility results under investigation here.

Let Q be the finitely axiomatized system of Robinson arithmetic. By $\underline{n}$ we mean the numeral in the language of arithmetic for the natural number n. (This numeral will be of the form $s \ldots s0$, with n occurrences of the symbol s for the successor function.) We assume that each syntactic item φ has its own unique code number $\#(\varphi)$. By $\overline{\varphi}$ we mean $\underline{\#(\varphi)}$, that is, the numeral for that code number. If $n = \#(\varphi)$ then of course $\varphi = \#^{-1}(n)$.

We mention here some basic results in the theory of the coding of syntax—our Theorems 1–3 below. That these three (meta)theorems can be proved using Core Logic $\mathbb{C}$ as one's *meta*logic, and with the deducibility sign $\vdash$ representing $\mathbb{C}$-deducibility in the *object* language, is demonstrated in [11].

THEOREM 1 (Representability of recursive functions). *For every* k*-place recursive function* $f(x_1, \ldots, x_k)$ *there is a formula* $\varphi(x_1, \ldots, x_k, y)$ *with just the indicated* $k + 1$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

variables free, such that for all natural numbers $n_1, \ldots, n_k$

$$\varphi(\underline{n_1}, \ldots, \underline{n_k}, a) \dashv_{\mathsf{Q}}\vdash a = \underline{f(n_1, \ldots, n_k)},$$

with the parameter a *not occurring in* $\varphi(x_1, \ldots, x_k, y)$.

THEOREM 2 (Strong representability of decidable relations).[7] *Every decidable relation* $R(n_1, \ldots, n_k)$ *on the natural numbers can be* strongly represented *by a formula* $\rho(x_1, \ldots, x_k)$ *in the following sense*:

> *for all natural numbers* $n_1, \ldots, n_k$,
> *if* $R(n_1, \ldots, n_k)$, *then* $\mathsf{Q} \vdash \rho(\underline{n_1}, \ldots, \underline{n_k})$; *and*
> *if not-*$R(n_1, \ldots, n_k)$, *then* $\mathsf{Q}, \rho(\underline{n_1}, \ldots, \underline{n_k}) \vdash \bot$.

THEOREM 3. *Suppose that* $\mathcal{S}$ *is a theory that interprets* Q. *Let* χ *be the sentence in the language of* $\mathcal{S}$ *that axiomatizes* Q *upon the interpretation in question. Then* $\mathcal{S}$ *provides* **fixed points**, *in the following sense*:

> *for every formula* $\psi(x, y_1, \ldots, y_n)$ *with just the indicated variables free, there is a formula* $\varphi(y_1, \ldots, y_n)$ *with just* $y_1, \ldots, y_n$ *free, such that*
>
> $$\chi, \psi(\overline{\varphi(t_1, \ldots, t_n)}, t_1, \ldots, t_n) \vdash \varphi(t_1, \ldots, t_n)$$
>
> *and*
>
> $$\chi, \varphi(t_1, \ldots, t_n) \vdash \psi(\overline{\varphi(t_1, \ldots, t_n), t_1, \ldots, t_n}).$$

Thus $\varphi(t_1, \ldots, t_n)$ *is interdeducible, modulo* χ, *with* $\psi(\overline{\varphi(t_1, \ldots, t_n)}, t_1, \ldots, t_n)$. *Put another way, the sentence* $\varphi(t_1, \ldots, t_n)$ *'says of itself'* (*via the coding, modulo* χ) *that it has the property* $\psi(_, t_1, \ldots, t_n)$.

Proof. See [6]; and see [11] for their reprisal in the constructive and relevant system $\mathbb{C}$ of Core Logic. □

4.1. Predicates for formal provability in a system. It is useful to talk more generally about systems $\mathcal{S}$ of formal arithmetic. Every system $\mathcal{S}$ that we shall consider will be assumed to be complete for bounded sentences: that is, if φ is a true sentence of the formal language of arithmetic containing no unrestricted quantifications, then $\mathcal{S}$ proves φ. Robinson's system Q has this property. (By 'true' here, of course, we mean true *in the intended model* $\mathbb{N}$ *of the natural numbers*.)

The two-place proof-predicate $\mathrm{Pf}_{\mathcal{S}}(x, y)$ in the formal language of first-order arithmetic is so defined that it strongly represents the relation 'x is the code number of a proof, in the system $\mathcal{S}$, of the sentence with code number y'. Such 'strong representation' takes place in the following sense:

> for all natural numbers n, m,
> if n *is* the code number of a proof, in the system $\mathcal{S}$, of the sentence with code number m, then $\mathcal{S}$ *proves* $\mathrm{Pf}_{\mathcal{S}}(\underline{n}, \underline{m})$; and
> if n is *not* the code number of a proof, in the system $\mathcal{S}$, of the sentence with code number m, then $\mathcal{S}$ *refutes* $\mathrm{Pf}_{\mathcal{S}}(\underline{n}, \underline{m})$.

[7] *Cf.* S. C. Kleene, *Introduction to Metamathematics*, Van Nostrand, 1950, p. 195. Kleene says that the relation is 'numeralwise expressible' rather than 'strongly representable'.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

The one-place provability-predicate $P_{\mathcal{S}}(y)$ is defined as $\exists x \mathrm{Pf}_{\mathcal{S}}(x, y)$. We shall re-write $P_{\mathcal{S}}(\overline{\varphi})$ as $P[\varphi]$. Sometimes we shall even omit the brackets, when no confusion can result.

§5. Exposition of Tarski's theorem. Tarski's theorem 'is' the Liar, in the formal setting introduced in §4. Let P be any unary formula. Once equipped with the rule below of 'P-Introduction' in a sufficiently strong system $\mathcal{S}$—thereby turning P into a *truth*-predicate—one could take the famous Liar sentence λ, which is the fixed point (in the sense of Theorem 3) of the unary formula $\neg P[x]$, and go through the reasoning of the Liar Paradox in order to show that $\mathcal{S}$ is inconsistent. This insight in the context of formal systems that express their own syntax is due to Tarski. The original insight, concerning natural languages, is due to Eubulides, over two millenia ago, as we have already pointed out.

THEOREM 4 (Tarski). *Suppose that the formal system $\mathcal{S}$ is sufficiently strong, and that it has a predicate P that satisfies the rules of P-Introduction and P-Elimination*:

$$\mathcal{S}, \theta \vdash P[\theta] \; ;^{8}$$

$$\mathcal{S}, P[\theta] \vdash \theta \; .$$

Then $\mathcal{S}$ is inconsistent.

Proof. Take λ as the fixed point of the unary formula $\neg P[x]$. So we have the fixed-point deducibilities

$$\mathcal{S}, \lambda \vdash \neg P[\lambda] \; ;$$

$$\mathcal{S}, \neg P[\lambda] \vdash \lambda \; .$$

Of course we also have (by main supposition) the λ-instances of the rules of P-Introduction and P-Elimination:

$$\mathcal{S}, \lambda \vdash P[\lambda] \; ;$$

$$\mathcal{S}, P[\lambda] \vdash \lambda \; .$$

Use the first fixed-point deducibility to construct the following proof Ω of $\mathcal{S}, \lambda, P[\lambda] \vdash \bot$:

$$\Omega \qquad \dfrac{\mathcal{S}, \lambda \vdash \neg P[\lambda] \quad \dfrac{P[\lambda] \vdash P[\lambda]}{P[\lambda], \neg P[\lambda] \vdash \bot}\,\neg\mathrm{L}}{\mathcal{S}, \lambda, P[\lambda] \vdash \bot}\,\mathrm{CUT.}$$

[8] Montague, in the hypotheses for his Lemma 3 (which is our Theorem 6), lays down a much weaker 'introductory' condition on P than P-Introduction. The weaker condition is his Condition 2:

$$\frac{\vdash \theta}{\mathcal{S} \vdash P[\theta],}$$

which says that $\mathcal{S}$ proves $P[\theta]$ (only) for *logical theorems* θ.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Then use Ω along with the second fixed-point deducibility and the λ-instance of the rule of P-Elimination to construct the following proof Θ of $\mathcal{S} \vdash \lambda$:

$$\dfrac{\Theta \qquad \dfrac{\dfrac{\dfrac{\Omega}{\mathcal{S}, P[\lambda] \vdash \lambda \qquad \mathcal{S}, \lambda, P[\lambda] \vdash \bot}\,\text{CUT}}{\mathcal{S}, P[\lambda] \vdash \bot}\,\neg\text{L}}{\mathcal{S} \vdash \neg P[\lambda]} \qquad \mathcal{S}, \neg P[\lambda] \vdash \lambda}{\mathcal{S} \vdash \lambda}\,\text{CUT.}$$

Finally use Θ and Ω once again, and the λ-instance of the rule of P-Introduction to construct the following proof of $\mathcal{S} \vdash \bot$:

$$\dfrac{\dfrac{\Theta}{\mathcal{S} \vdash \lambda} \qquad \dfrac{\dfrac{\Omega}{\mathcal{S}, \lambda \vdash P[\lambda] \qquad \mathcal{S}, \lambda, P[\lambda] \vdash \bot}\,\text{CUT}}{\mathcal{S}, \lambda \vdash \bot}}{\mathcal{S} \vdash \bot}\,\text{CUT.}$$

□

We remind the reader that this result, and the reasoning that establishes it, is on the 'arithmetical theorizing' side of the topical divide essayed upon above. That is to say, it shows that there is no definiens P in the arithmetical language that can be made subject to the conditions expressed by the rules of P-Introduction and P-Elimination. For, *if* there were such a definiens P, it would engender the *in*consistency of the formal system $\mathcal{S}$ of arithmetic, which—*ex hypothesi*, or in light of metamathematical proof for particular choices of $\mathcal{S}$—is sufficiently strong.

§6. Exposition of a theorem of Löb. The following are well-known facts about the provability predicate. They can be shown to hold of any defined provability predicate P obtained *via* the method of Gödel coding of sentences and proofs. We continue to assume that $\mathcal{S}$ is a consistent and sufficiently strong system of first-order arithmetic.

1. $\dfrac{\mathcal{S} \vdash \varphi}{\mathcal{S} \vdash P\varphi.}$
2. $\mathcal{S}, P\varphi, P[\varphi \to \psi] \vdash P\psi$.[9]
3. $\mathcal{S}, P\varphi \vdash P[P\varphi]$.[10]

LEMMA 1. $\dfrac{\mathcal{S}, \varphi \vdash \psi}{\mathcal{S}, P\varphi \vdash P\psi}$

[9] See Boolos, *The Logic of Provability*, Cambridge University Press, 1993, for proofs, terminating on p. 44.

[10] Boolos (op. cit., pp. 46–49) proves, for the case where the system $\mathcal{S}$ is (or interprets) Peano Arithmetic, a more general result than (3), from which (3) immediately follows: for Σ_1-sentences φ, we have $PA, \varphi \vdash P[\varphi]$. Since $P[\varphi]$ is Σ_1, taking $P[\varphi]$ as φ yields Condition (3) (for $\mathcal{S} = PA$). We do not, however, need to resort to having $\mathcal{S}$ so strong that it interprets PA. It suffices if $\mathcal{S}$ interprets, in addition to Q, the axiom scheme of Σ_1-induction. So for $\mathcal{S}$ here we can take a system that interprets $I\Sigma_1$ (I owe this observation to Vann McGee.) For more on the systems $I\Sigma_n$, see [1].

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Proof.
$$\dfrac{\dfrac{\dfrac{\mathcal{S},\varphi\vdash\psi}{\mathcal{S}\vdash\varphi\to\psi}\to\text{R}}{\mathcal{S}\vdash P[\varphi\to\psi]}(1) \qquad \begin{matrix}(2):\\ \mathcal{S},P\varphi,P[\varphi\to\psi]\vdash P\psi\end{matrix}}{\mathcal{S},P\varphi\vdash P\psi}\ \text{CUT.}$$
□

LEMMA 2. $\dfrac{\mathcal{S},P\varphi\vdash P[P[\varphi]\to\psi]}{\mathcal{S},P\varphi\vdash P\psi}$.

Proof.
$$\dfrac{\begin{matrix}(3):\\ \mathcal{S},P\varphi\vdash P[P\varphi]\end{matrix} \quad \dfrac{\begin{matrix}\text{Hyp:}\\ \mathcal{S},P\varphi\vdash P[P[\varphi]\to\psi]\end{matrix} \quad \begin{matrix}(2):\\ \mathcal{S},P[P\varphi],P[P[\varphi]\to\psi]\vdash P\psi\end{matrix}}{\mathcal{S},P\varphi,P[P\varphi]\vdash P\psi}\ \text{CUT}}{\mathcal{S},P\varphi\vdash P\psi.}\ \text{CUT}$$
□

LEMMA 3. $\dfrac{\mathcal{S},\varphi\vdash P[\varphi]\to\psi}{\mathcal{S},P\varphi\vdash P\psi.}$

Proof.
$$\dfrac{\dfrac{\mathcal{S},\varphi\vdash P[\varphi]\to\psi}{\mathcal{S},P\varphi\vdash P[P[\varphi]\to\psi]}(L1)}{\mathcal{S},P\varphi\vdash P\psi.}(L2)$$
□

LEMMA 4. $\dfrac{\mathcal{S},\psi\vdash P[\psi]\to\varphi \qquad \mathcal{S},P\varphi\vdash\varphi}{\mathcal{S},P\psi\vdash\varphi.}$

Proof.
$$\dfrac{\dfrac{\mathcal{S},\psi\vdash P[\psi]\to\varphi}{\mathcal{S},P\psi\vdash P\varphi}(L3) \qquad \mathcal{S},P\varphi\vdash\varphi}{\mathcal{S},P\psi\vdash\varphi}\ \text{CUT.}$$
□

LEMMA 5. $\dfrac{\mathcal{S},\psi\vdash P[\psi]\to\varphi \qquad \mathcal{S},P[\psi]\to\varphi\vdash\psi \qquad \mathcal{S},P\varphi\vdash\varphi.}{\mathcal{S}\vdash\varphi}$

Proof.
$$\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{S},\psi\vdash P[\psi]\to\varphi \quad \mathcal{S},P\varphi\vdash\varphi}{\mathcal{S},P\psi\vdash\varphi}(L4)}{\mathcal{S}\vdash P[\psi]\to\varphi}\to\text{R} \qquad \mathcal{S},P[\psi]\to\varphi\vdash\psi}{\mathcal{S}\vdash\psi}\ \text{CUT} \qquad \dfrac{\mathcal{S},\psi\vdash P[\psi]\to\varphi \quad \mathcal{S},P\varphi\vdash\varphi}{\mathcal{S},P\psi\vdash\varphi}(L4)}{\mathcal{S}\vdash P\psi}(1)}{\mathcal{S}\vdash\varphi}\ \text{CUT.}$$
□

THEOREM 5 (Löb).[11] *For all* φ, *if* $\mathcal{S},P\varphi\vdash\varphi$ *then* $\mathcal{S}\vdash\varphi$.

Proof. Let φ be any sentence. Consider the predicate $P[x]\to\varphi$. It has a fixed point. Call it ψ. Thus we have $\mathcal{S},\psi\vdash P[\psi]\to\varphi$ and $\mathcal{S},P[\psi]\to\varphi\vdash\psi$. Suppose now that $\mathcal{S},P\varphi\vdash\varphi$. Then by Lemma 5 it follows that $\mathcal{S}\vdash\varphi$. □

[11] See [3].

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

COROLLARY 1 (Gödel's Second). *Suppose the logic of $\mathcal{S}$ contains the Absurdity Rule (from $\perp$ infer any sentence you like). Suppose $\mathcal{S}$ is consistent. Then $\mathcal{S}$ does not refute $P[0 = 1]$.*

Proof. Suppose for *reductio* that $\mathcal{S}, P[0 = 1] \vdash \perp$. Then by the Absurdity Rule we have $\mathcal{S}, P[0 = 1] \vdash 0 = 1$. It follows by Theorem 5 that $\mathcal{S} \vdash 0 = 1$. But $\mathcal{S}$ interprets Q, and $\mathsf{Q}, 0 = 1 \vdash \perp$. So $\mathcal{S} \vdash \perp$, contrary to the assumed consistency of $\mathcal{S}$. Thus $\mathcal{S}$ does not refute $P[0 = 1]$. □

Comments. In this section, we have assumed only such conditions on the provability predicate P as can be shown to hold of any defined provability predicate P obtained *via* the method of Gödel coding of sentences and proofs in any consistent extension of Q. In the next section we do something slightly different. We consider various conditions that one might impose axiomatically on a *primitive* (not: defined) 'provability' predicate P. P may be thought of as a single primitive monadic predicate. We call it a 'provability' predicate (note the scare quotes), and keep using the suggestive letter 'P', for the sake of continuity. But it should be borne in mind that the predicate P might in some cases be made subject to conditions weaker than those satisfied by an actual (defined) provability predicate, or indeed in some cases stronger—more like a *truth* predicate than a provability predicate.Since P is being taken as a primitive, it would acquire its own Gödel code number. Complex expressions involving P (including formal proofs involving sentences containing P) would thereby be assigned their own code numbers according to the usual coding method. Since we shall be considering only theories that interpret Robinson arithmetic Q, we shall be able to appeal to the fixed-point theorem for those theories (see Theorem 3). And that theorem of course applies to all sentences of the theory in question, including sentences that contain P.

Some of the conditions to be imposed on P will be identical to, or closely related to, Conditions 1–3 above, which provably hold (as already remarked) when P is a genuine (defined) provability predicate in a strong enough system. But the reader must bear in mind that in what follows we are merely assuming that certain conditions hold of P, and investigating the consequences of such assumptions.

§7. Exposition of a theorem of Montague. First, a simple result in Intuitionistic Logic. We shall invoke it in due course in our constructive proof of Lemma 7.

LEMMA 6. $\dfrac{\rho, \sigma \vdash \tau}{\vdash (\tau \to (\rho \to \neg\sigma)) \to (\rho \to \neg\sigma).}$

Proof. Using the parallelized form of →-Elimination, the proof is as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overbrace{\rho \quad , \quad \sigma}^{(4)} \quad \vdots}{\tau}}{\tau \to (\rho \to \neg\sigma)}(5) \qquad \dfrac{\dfrac{\dfrac{\rho \to \neg\sigma \quad \rho}{\perp}(1)}{\perp}(2) \qquad \dfrac{\neg\sigma \quad \sigma}{\perp}(1,3)}{\perp}(2,4)}{\perp}(3)}{\neg\sigma}(4)}{\rho \to \neg\sigma}(5)}{(\tau \to (\rho \to \neg\sigma)) \to (\rho \to \neg\sigma).}$$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

The sequent proof (in the sequent calculus for Intuitionistic Logic) that is isomorphic to this last natural deduction (whose eliminations are parallelized) is as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\sigma \;:\; \sigma}{\rho \;:\; \rho \qquad \sigma,\ \neg\sigma \;:}}{\rho,\ \sigma \;:\; \tau \qquad \rho,\ \sigma,\ \rho{\to}\neg\sigma \;:}}{\rho,\ \sigma,\ \tau \to (\rho \to \neg\sigma) \;:}}{\rho, \tau \to (\rho \to \neg\sigma) \;:\; \neg\sigma}}{\tau \to (\rho \to \neg\sigma) \;:\; \rho \to \neg\sigma}}{:\; (\tau \to (\rho \to \neg\sigma)) \to (\rho \to \neg\sigma).}$$

□

Note that we have just used empty space on the left of the colon (when the antecedent of the sequent in question is empty) and likewise empty space on the right of the colon, in place of $\bot$.

Consider now the following possible conditions on the provability predicate P for a formal system $\mathcal{S}$.[12] Condition 3 governing the conditional is weaker than the 'corresponding' Condition 2 on conditionals that was used for Löb's theorem in §6.

1. $\mathcal{S} \vdash P[P[\theta] \to \theta]$.
2. $\dfrac{\vdash \theta}{\mathcal{S} \vdash P[\theta].}$
3. $\dfrac{\mathcal{S} \vdash P[\varphi \to \psi] \quad \mathcal{S} \vdash P[\varphi]}{\mathcal{S} \vdash P[\psi].}$
4. $\mathcal{S}, P[\theta] \vdash \theta$.

LEMMA 7. $\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\mathcal{S} \vdash P[\chi \to \neg\varphi].}$

Proof.

$$\dfrac{\dfrac{\dfrac{\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\vdash (P[\chi \to \neg\varphi] \to (\chi \to \neg\varphi)) \to (\chi \to \neg\varphi)}(L6)}{\mathcal{S} \vdash P[(P[\chi \to \neg\varphi] \to (\chi \to \neg\varphi)) \to (\chi \to \neg\varphi)]}(2) \qquad \dfrac{(1)\;:}{\mathcal{S} \vdash P[P[\theta] \to \theta]}}{\mathcal{S} \vdash P[\chi \to \neg\varphi].}(3)$$

□

LEMMA 8. $\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi] \quad \mathcal{S}, P[\chi \to \neg\varphi] \vdash \varphi}{\mathcal{S} \vdash \varphi.}$

Proof. $\dfrac{\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\mathcal{S} \vdash P[\chi \to \neg\varphi]}(L7) \qquad \mathcal{S}, P[\chi \to \neg\varphi] \vdash \varphi}{\mathcal{S} \vdash \varphi}\ \text{CUT.}$ □

LEMMA 9. $\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\mathcal{S} \vdash \chi \to \neg\varphi.}$

[12] From now on we drop the scare quotes around 'provability'.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Proof. $$\dfrac{\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\mathcal{S} \vdash P[\chi \to \neg\varphi]}(L7) \qquad \dfrac{\overset{(4):}{\mathcal{S}, P[\theta] \vdash \theta}}{\mathcal{S}\,,\, P[\chi \to \neg\varphi] \vdash \chi \to \neg\varphi}}{\mathcal{S} \vdash \chi \to \neg\varphi}\ \text{CUT.}$$ □

OBSERVATION 1. *Suppose that $\mathcal{S}$ is a theory that interprets* Q. *Let χ be the sentence in the language of $\mathcal{S}$ that axiomatizes* Q *upon the interpretation in question. Then $\mathcal{S} \vdash \chi$.*

THEOREM 6 (Montague).[13] *Any theory $\mathcal{S}$ that interprets* Q *and whose provability predicate P satisfies conditions* (1)–(4) *is inconsistent.*

Proof. Let χ be the sentence in the language of $\mathcal{S}$ that axiomatizes Q upon the interpretation in question. By Observation 1 we have

$$\mathcal{S} \vdash \chi.$$

Consider the unary formula $P[\chi \to \neg x]$. By Theorem 3 this formula has a fixed point. Call it φ. So we have both

$$\chi, P[\chi \to \neg\varphi] \vdash \varphi$$

and

$$\chi, \varphi \vdash P[\chi \to \neg\varphi].$$

From the last three displayed deducibilities we now reason as follows:

$$\dfrac{\dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi] \qquad \dfrac{\mathcal{S} \vdash \chi \quad \chi, P[\chi \to \neg\varphi] \vdash \varphi}{\mathcal{S}, P[\chi \to \neg\varphi] \vdash \varphi}\text{CUT}}{\mathcal{S} \vdash \varphi}(L8) \qquad \dfrac{\mathcal{S} \vdash \chi \qquad \dfrac{\chi, \varphi \vdash P[\chi \to \neg\varphi]}{\mathcal{S} \vdash \chi \to \neg\varphi}(\text{L9})}{\mathcal{S} \vdash \neg\varphi}\to\text{-E}}{\mathcal{S} \vdash \bot}\ \neg\text{-E.}$$

□

Supplying the proofs of Lemmas 8 and 9, and the proof of Lemma 7 within that of Lemma 9, we obtain the following proof (so wide that it is in landscape mode). Within the proof of Lemma 7 we have got rid of the coloration and repetition of sentences. The premise of Lemma 7 is now written in red, and its conclusion in blue, in order to help the reader to see what proof-work would need to be substituted for the vertical dots.

[13] This is Montague's Lemma 3 in [5] at p. 289. We are rendering Montague's proof more perspicuous by casting it into natural deduction in the metalanguage.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

$$
\dfrac{
\dfrac{
\begin{array}{c} \chi, \varphi \vdash P[\chi \rightarrow \neg\varphi] \\ \vdots \\ \mathcal{S} \vdash P[\chi \rightarrow \neg\varphi] \end{array}
\qquad
\dfrac{\mathcal{S} \vdash \chi \quad \chi, P[\chi \rightarrow \neg\varphi] \vdash \varphi}{\mathcal{S}, P[\chi \rightarrow \neg\varphi] \vdash \varphi}\,\text{CUT}
}{\mathcal{S} \vdash \varphi}\,\text{CUT}
\qquad
\dfrac{
\mathcal{S} \vdash \chi
\qquad
\dfrac{
\dfrac{
\dfrac{
\dfrac{\chi, \varphi \vdash P[\chi \rightarrow \neg\varphi]}{\vdash (P[\chi \rightarrow \neg\varphi] \rightarrow (\chi \rightarrow \neg\varphi)) \rightarrow (\chi \rightarrow \neg\varphi)}\,(L6)
}{\mathcal{S} \vdash P[(P[\chi \rightarrow \neg\varphi] \rightarrow (\chi \rightarrow \neg\varphi)) \rightarrow (\chi \rightarrow \neg\varphi)]}\,(2)
\quad
\dfrac{\begin{array}{c}(1): \\ \mathcal{S} \vdash [P[\theta] \rightarrow \theta]\end{array}}{\mathcal{S} \vdash P[P[\chi \rightarrow \neg\varphi] \rightarrow (\chi \rightarrow \neg\varphi)]}
}{\mathcal{S} \vdash P[\chi \rightarrow \neg\varphi]}\,(3)
\quad
\dfrac{\begin{array}{c}(4): \\ \mathcal{S}, P[\theta] \vdash \theta\end{array}}{\mathcal{S}, P[\chi \rightarrow \neg\varphi] \vdash \chi \rightarrow \neg\varphi}
}{\mathcal{S} \vdash \chi \rightarrow \neg\varphi}\,\text{CUT}
}{\mathcal{S} \vdash \neg\varphi}\,\rightarrow\text{-E}
}{\mathcal{S} \vdash \bot}\,\neg\text{-E}
$$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

7.1. Comments on Montague's conditions on a formal provability predicate. The conditions on a formal provability predicate that Montague's Theorem shows cannot be met by any sufficiently strong, consistent theory are, in one regard at least, rather weak. Note that Condition 4

$$\mathcal{S}, P[\theta] \vdash \theta$$

can be thought of as the rule of P-Elimination. Now of course the straightforwardly corresponding rule of P-Introduction

$$P\text{-Introduction} \qquad \mathcal{S}, \theta \vdash P[\theta]$$

would turn P into a truth-predicate. This rule says that for any θ, the system $\mathcal{S}$ proves $P[\theta]$ from the assumption θ. This is obviously too strong a condition to impose on provability. Nevertheless, we shall continue to make merely formal use of a predicate P, so that the results produced by working with the overly strong rule will be easy to compare with earlier results above.

LEMMA 10. *P-Introduction, along with P-Elimination (Condition 4), implies that P satisfies Conditions 1–3.*

Proof. Condition 1 is derived as follows:

$$\dfrac{\dfrac{(4):\ \mathcal{S}, P[\theta] \vdash \theta}{\mathcal{S} \vdash P[\theta]\to\theta}\ {\to}\mathrm{R} \qquad P\text{-Intro:}\ \mathcal{S}, P[\theta]\to\theta \vdash P[P[\theta]\to\theta]}{\mathcal{S} \vdash P[P[\theta]\to\theta]}\ \text{CUT.}$$

Condition 2 is derived as follows:

$$\dfrac{\vdash \theta \quad P\text{-Intro:}\ \mathcal{S}, \theta \vdash P[\theta]}{\mathcal{S} \vdash P[\theta]}\ \text{CUT.}$$

Condition 3 is derived as follows:

$$\dfrac{P\text{-Intro}: \dfrac{\dfrac{\dfrac{}{\varphi \vdash \varphi \quad \psi \vdash \psi}\ {\to}\text{-L}}{\varphi\to\psi, \varphi \vdash \psi}\quad \dfrac{\mathcal{S} \vdash P[\varphi \to \psi] \quad (4):\ \mathcal{S}, P[\varphi\to\psi] \vdash \varphi\to\psi}{\mathcal{S} \vdash \varphi\to\psi}\ \text{CUT}}{\mathcal{S}, \varphi \vdash \psi}\ \text{CUT} \qquad \dfrac{(4):\ \mathcal{S} \vdash P[\varphi]\ \mathcal{S}, P[\varphi] \vdash \varphi}{\mathcal{S} \vdash \varphi}\ \text{CUT}}{\mathcal{S} \vdash \psi}\ \text{CUT.}$$

$$\mathcal{S}, \psi \vdash P[\psi] \qquad\qquad \mathcal{S} \vdash P[\psi]$$

□

§8. Exposition of McGee's version of a Liar paradox. Vann McGee, in [4], at p. 26, states what he calls 'Theorem 1.4 (Montague)'. This is Montague's Theorem 3 at p. 293 of [5]. It is not Montague's Lemma 3, p. 289 (our Theorem 6), nor is it Montague's Theorem 1, *loc. cit.*, pp. 292–293. Whereas Montague's Theorem 3 follows (by Lemma 10) as an immediate consequence of his Lemma 3 (which uses Conditions 1–3 instead of P-Introduction), McGee furnishes a detailed proof of Montague's

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Theorem 3 from simpler P-conditions, of which there are only two (see Theorem 7 below).[14]

McGee uses the introductory condition that the system $\mathcal{S}$ proves $P[\theta]$ whenever $\mathcal{S}$ proves θ. This condition is stronger than Condition 2 for Montague's Lemma 3 (that the system proves $P[\theta]$ whenever θ is a logical theorem). Yet it is not so strong as to make P into a truth-predicate. Rather, McGee's introductory condition is one that it would be plausible to think is satisfied by a provability predicate. Here are McGee's 'introductory' and 'eliminative' conditions for P:

$$\frac{\mathcal{S} \vdash \theta}{\mathcal{S} \vdash P[\theta]} \text{ (stronger than } \frac{\vdash \theta}{\mathcal{S} \vdash P[\theta]} \text{ , but plausible for provability);}$$

$$\mathcal{S}, P[\theta] \vdash \theta.$$

It may be a little surprising to the reader to learn that the Liar Paradox affects any sufficiently strong system $\mathcal{S}$ for which these P-conditions hold.

THEOREM 7 (McGee). *Suppose $\mathcal{S}$ is sufficiently strong, and that its provability predicate P satisfies McGee's introductory and eliminative conditions*

$$\frac{\mathcal{S} \vdash \theta}{\mathcal{S} \vdash P[\theta]} \quad \textit{and} \quad \mathcal{S}, P[\theta] \vdash \theta.$$

Then $\mathcal{S}$ is inconsistent.

Proof. Once again take λ as the fixed point of the unary formula $\neg P[x]$. So we have the deducibilities

$$\mathcal{S}, \lambda \vdash \neg P[\lambda] \text{ and } \mathcal{S}, \neg P[\lambda] \vdash \lambda\,.$$

Of course we also have the λ-instances of McGee's introductory and eliminative conditions on P:

$$\frac{\mathcal{S} \vdash \lambda}{\mathcal{S} \vdash P[\lambda]} \text{ and } \mathcal{S}, P[\lambda] \vdash \lambda.$$

Here is a faithful formalization of McGee's reasoning, using the last four displayed conditions. First, construct the following one-step sequent proof Π of $\mathcal{S} \vdash \lambda$:

$$\Pi \qquad \frac{\mathcal{S}, \neg P[\lambda] \vdash \lambda \qquad \mathcal{S}, P[\lambda] \vdash \lambda}{\mathcal{S} \vdash \lambda.}$$

[14] Being no expert on the exact historical order and possibility of disagreements about priority in these complicated matters, the present author feels obliged to share with his reader the following helpful remark from an anonymous referee:

> The paradox listed as McGee's version of the Liar paradox (§8) is actually what I had been taught by the folklore and by my teachers (informally at least) to call the Montague paradox, and not the more complicated construction in §7.

The present author is confining his efforts to the constructive formalization of the deductive reasoning he finds in the various papers cited, and making attributions accordingly, trusting that the various authors cited would not have overlooked any need to accord credit for any their results to predecessors.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Then use Π to reason as follows:

$$\dfrac{\Pi\quad \mathcal{S}\vdash\lambda \qquad \dfrac{\dfrac{\mathcal{S}\vdash\lambda}{\mathcal{S}\vdash P[\lambda]}\quad \dfrac{\mathcal{S},\lambda\vdash\neg P[\lambda] \qquad \dfrac{P[\lambda]\vdash P[\lambda]}{P[\lambda],\neg P[\lambda]\vdash\bot}\,\text{CUT}}{\mathcal{S},P[\lambda],\lambda\vdash\bot}\,\text{CUT}}{\mathcal{S},\lambda\vdash\bot}\,\text{CUT}.}{\mathcal{S}\vdash\bot}$$

This ends our formalization of McGee's reasoning. □

McGee's reasoning looks strictly classical, since there is a *prima facie* application of classical dilemma in the subproof Π.[15] But this classicality is only apparent; the reasoning can be constructivized. To see this, let us give the name Σ to the following (constructive) subproof at the top right of the immediately foregoing proof-display:

$$\Sigma\qquad \dfrac{\mathcal{S},\lambda\vdash\neg P[\lambda] \qquad \dfrac{P[\lambda]\vdash P[\lambda]}{P[\lambda],\neg P[\lambda]\vdash\bot}\,\text{CUT}}{\mathcal{S},P[\lambda],\lambda\vdash\bot}\,\text{CUT}.$$

Then continue constructively as follows:

$$\dfrac{\dfrac{\Sigma}{\mathcal{S},P[\lambda]\vdash\lambda \quad \mathcal{S},P[\lambda],\lambda\vdash\bot}\,\text{CUT}}{\dfrac{\dfrac{\mathcal{S},P[\lambda]\vdash\bot}{\mathcal{S}\vdash\neg P[\lambda]}\,\neg\text{R} \qquad \mathcal{S},\neg P[\lambda]\vdash\lambda}{\mathcal{S}\vdash\lambda}\,\text{CUT}.}$$

Call this last proof Π^*. Now finish off (constructively) as follows:

$$\dfrac{\Pi^*\quad \mathcal{S}\vdash\lambda \qquad \dfrac{\dfrac{\mathcal{S}\vdash\lambda}{\mathcal{S}\vdash P[\lambda]}\quad \dfrac{\Sigma}{\mathcal{S},P[\lambda],\lambda\vdash\bot}\,\text{CUT}}{\mathcal{S},\lambda\vdash\bot}\,\text{CUT}}{\mathcal{S}\vdash\bot}\,\text{CUT}.$$

§9. Exposition of a theorem of McGee about ω-inconsistency. If English had a suitably short and stylistically attractive word for 'lengthy exposition', it would have been the first word in the title of this section. It is necessary to go through the logical details that lie ahead in this section with the same granular precision as was involved in §5–§8. We thereby set ourselves the strongest challenge in the proposed project of showing (on the *other* side of the topical divide) that the reasoning with a *primitive* predicate would be so formalizable as to be revealed as logico-semantically paradoxical.

DEFINITION 1. *$\mathcal{S}$ is* ω-inconsistent *just in case for some unary formula ψx, we have that*

$$\textit{for every natural number } n,\ \mathcal{S}\vdash\psi\underline{n},$$

and yet have also that

$$\mathcal{S},\forall x(Nx\rightarrow\psi x)\vdash\bot.$$

[15] This corresponds to McGee's step to his (iii)—*ibid.*, at p. 26.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

McGee proves another result, this time establishing the ω-inconsistency of any system $\mathcal{S}$ that interprets Q,[16] and that satisfies certain conditions on a 'provability' predicate P (primitive or defined).

The conditions McGee places on $\mathcal{S}$ are as follows. Note that Conditions 1 and 2 are as in the foregoing discussion of Löb's Theorem.

1. $\dfrac{\mathcal{S} \vdash \varphi}{\mathcal{S} \vdash P\varphi.}$
2. $\mathcal{S}, P\varphi, P[\varphi \to \psi] \vdash P\psi.$
3. $\mathcal{S}, P[\neg\varphi], P[\varphi] \vdash \bot.$
4. $\mathcal{S}, \forall x(Nx \to P[\varphi\underline{x}]) \vdash P[\forall x(Nx \to \varphi x)].$

9.1. Introductory and eliminative aspects of P. The reader will recall our having noted that McGee's Condition 2 here (also used by Löb), dealing with P and $\to$, is stronger than Montague's corresponding Condition 3. Neither Löb nor McGee, however, have a reasonably full-blooded rule of P-Elimination like Montague's Condition 4. One needs a certain amount of 'P-Introduction' and a certain amount of 'P-Elimination' in order to cross the line into the territory of semantic paradox. Löb has very little of the latter—indeed, only his Condition 2 (which he shares with McGee) has any hint of 'eliminativeness' to it. On the 'introductory' side, Montague has only the rather weak Condition 2, whereas McGee now has the significantly stronger Condition 1 (which he shares with Löb).

It would appear, then, that the untoward result of ω-inconsistency that McGee establishes (see Theorem 8 below) from the four conditions just listed above must engender, via its Conditions 2–4 reprised here:

2. $\mathcal{S}, P\varphi, P[\varphi \to \psi] \vdash P\psi.$
3. $\mathcal{S}, P[\neg\varphi], P[\varphi] \vdash \bot.$
4. $\mathcal{S}, \forall x(Nx \to P[\varphi\underline{x}]) \vdash P[\forall x(Nx \to \varphi x)].$

just enough 'eliminativeness' regarding P, to match the introductory flavor of Condition 1. That eliminativeness can be detected clearly in Conditions 2 and 3. Both these conditions invite one to think of P as being 'locally' eliminable—'locally', because the contexts involve the conditional and negation, respectively. Let us illustrate this thought by taking each of Conditions 2 and 3 in turn.

First, consider Condition 2. Within $\mathcal{S}$, the obvious reasoning that would be invited in order to 'derive' Condition 2 would involve eliminating P from $P\varphi$ to get φ, then eliminating P from $P[\varphi \to \psi]$ to get $\varphi \to \psi$, then performing $\to$-Elimination to get ψ, and finally appealing to Condition 1 to reintroduce P so as to get $P\psi$.

Secondly, consider Condition 3. Within $\mathcal{S}$, the obvious reasoning that would be invited in order to 'derive' Condition 3 would involve eliminating P from $P\varphi$ to get φ, then eliminating P from $P[\neg\varphi]$ to get $\neg\varphi$, then performing $\neg$-Elimination so as to get $\bot$.

McGee's Condition 4 invites a similar line of reflection. Note that this is the only condition (among those of all three of the writers under consideration in this

[16] Strictly speaking, McGee uses the weaker, infinitely axiomatized Robinson arithmetic R (supplemented by the assertions that the successor function is one–one and that zero is not a successor) in place of the stronger, finitely axiomatized Robinson arithmetic Q. Nothing in our exposition turns on this difference.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

study) that involves the proper understanding of universal (numerical) quantification. Interestingly, however, Condition 4 contributes—or so the present author believes—a considerable degree of both 'introductory' and 'eliminative' force for P within $\mathcal{S}$. The matter is quite subtle, and merits further explanation.

We have just considered the natural lines of reasoning 'within $\mathcal{S}$' that might generate Conditions 2 and 3. So let us now embark on a similar thought-experiment for Condition 4.

The natural way to 'force' a universal numerical quantification to be true if all its numerical instances are true is to adopt the infinitary ω-rule

$$\frac{\varphi\underline{0} \quad \varphi\underline{1} \quad \ldots \quad \varphi\underline{n} \quad \ldots}{\forall x(Nx \rightarrow \varphi x).}$$

Suppose that $\forall x(Nx \rightarrow P[\varphi\underline{x}])$. For each natural number n, the system $\mathcal{S}$ proves $N\underline{n}$. Hence by $\forall$-Elimination and $\rightarrow$-Elimination we can infer $P[\varphi\underline{n}]$. Partial progress report: we now have, for each natural number n, that

$$\mathcal{S}\ ,\ \forall x(Nx \rightarrow P[\varphi\underline{x}])\ \vdash\ P[\varphi\underline{n}].$$

Put another way: we have, for each natural number n, a proof of the form

$$\begin{array}{c} \underbrace{\mathcal{S}\ ,\ \forall x(Nx \rightarrow P[\varphi\underline{x}])} \\ \vdots \\ P[\varphi\underline{n}]. \end{array}$$

At this stage a natural-seeming eliminative move would seem to be invited, to infer $\varphi\underline{n}$. Note, however, that this move would be made (within the system $\mathcal{S}$) under the supposition $\forall x(Nx \rightarrow P[\varphi\underline{x}])$. So the P-Elimination rule being applied here would take the considerably more powerful form

$$\frac{\mathcal{S}, \Delta \vdash P[\theta]}{\mathcal{S}, \Delta \vdash \theta} \qquad \text{or} \qquad \begin{array}{c} \underbrace{\mathcal{S}\ ,\ \Delta} \\ \vdots \\ \dfrac{P[\theta]}{\theta,} \end{array}$$

which is nothing less than the straightforward rule of P-Elimination that can be stated simply as

$$\frac{P[\theta]}{\theta.}$$

Imagine, however, that we stifle our scruples about using such a powerful rule, and proceed in our quest to exploit the ω-rule in order to attain our ultimate goal. Application of the ω-rule at this stage yields only $\forall x(Nx \rightarrow \varphi x)$. The picture is as follows:

$$\frac{\left\{\begin{array}{c} \underbrace{\mathcal{S}\ ,\ \forall x(Nx \rightarrow P[\varphi\underline{x}])} \\ \vdots \\ P[\varphi\underline{n}] \end{array}\right\}_{n\in\omega}}{\forall x(Nx \rightarrow \varphi x).}$$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Our ultimate goal, however, is $P[\forall x(Nx \to \varphi x)]$. In order to attain it, we appear to need brute force: simply infer it now from $\forall x(Nx \to \varphi x)$. This would require a rule of P-Introduction in the form

$$\frac{\underbrace{\mathcal{S}, \Delta}}{\begin{array}{c}\vdots \\ \theta \end{array}}{P[\theta],}$$

which simplifies down to

$$\frac{\theta}{P[\theta]}\ .$$

Of course, the imputation that these straightforward introduction and elimination rules for P are what is *required* to do the job has to be taken *cum grano salis*, since Δ is far more general than $\mathcal{S}, \forall x(Nx \to \varphi x)$. But the impression is inescapable that Condition 4 embodies a requirement of 'introductoriness' and 'eliminativeness' on the part of P to a degree that makes it a candidate for (at least some kind of) semantic paradoxicality. And that is what McGee's theorem shows—at least, for one who contends that there is this organic connection between:

(i) the arithmetical impossibility and (ii) the logico-semantic paradoxicality,

that arise, respectively, upon

(i) assuming the predicate to be definable, or (ii) taking it as a primitive.

9.2. Back to the proof of McGee's theorem.

THEOREM 8 (McGee). *Suppose that $\mathcal{S}$ interprets* Q *and satisfies McGee's Conditions* 1–4 *stated above. Then $\mathcal{S}$ is ω-inconsistent.*

Proof. To follow in due course. Our overriding concern here is to ensure that the proof is constructive. □

LEMMA 11. $$\frac{\forall x(Nx \to \Psi x)}{\forall x(Nx \to \Psi sx).}$$

Proof.
$$\dfrac{\dfrac{\dfrac{\overset{(1)}{\overline{Na}}}{Nsa} \quad \dfrac{\forall x(Nx \to \Psi x)}{Nsa \to \Psi sa}}{\Psi sa}}{\dfrac{Na \to \Psi sa}{\forall x(Nx \to \Psi sx).}}\ {}^{(1)}$$
□

LEMMA 12. $$\frac{\forall x(Nx \to \forall z(\Xi xz \to \Theta z)) \quad \forall x(Nx \to \Xi xt)}{\forall x(Nx \to \Theta t).}$$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Proof.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{Na}(1) \quad \dfrac{\forall x(Nx \to \forall z(\Xi xz \to \Theta z))}{Na \to \forall z(\Xi az \to \Theta z)}}{\forall z(\Xi az \to \Theta t)}}{\Xi at \to \Theta t} \quad \dfrac{\dfrac{}{Na}(1) \quad \dfrac{\forall x(Nx \to \Xi xt)}{Na \to \Xi at}}{\Xi at}}{\Theta t}}{Na \to \Theta t}(1)}{\forall x(Nx \to \Theta t).}$$

□

9.3. McGee's diagonalization. Take $G(x, y, z, w_0)$ to abbreviate the formula that formalizes the following open sentence:[17]

$$w_0 \text{ is a formula of the form}$$

$$H(x, y, z) \wedge ((x = 0 \wedge z = y) \vee \exists w(Nw \wedge x = sw \wedge z = \overline{\forall v(\#^{-1}(w_0)^{xz}_{wv} \to \tau(v))})).$$

Now use Theorem 3 to find a formula $F(x, y, z)$ so that $\mathcal{S}$ proves

$$\forall x \forall y \forall z(F(x, y, z) \leftrightarrow G(x, y, z, \overline{F(x, y, z)}))$$

—that is, so that $\mathcal{S}$ proves

$$\forall x \forall y \forall z(F(x, y, z) \leftrightarrow$$

$$[\overline{F(x, y, z)} \text{ is a formula of the form}$$

$$H(x, y, z) \wedge ((x = 0 \wedge z = y) \vee \exists w(Nw \wedge x = sw \wedge z = \overline{\forall v(\#^{-1}(\overline{F(x, y, z)})^{xz}_{wv} \to \tau(v))})))].$$

Note that

$$\#^{-1}(\overline{F(x, y, z)})^{xz}_{wv} = F(x, y, z)^{xz}_{wv} = F(w, y, v).$$

So $\mathcal{S}$ proves

$$\forall x \forall y \forall z(F(x, y, z) \leftrightarrow$$

$$[\overline{F(x, y, z)} \text{ is a formula of the form}$$

$$H(x, y, z) \wedge ((x = 0 \wedge z = y) \vee \exists w(Nw \wedge x = sw \wedge z = \overline{\forall v(F(w, y, v) \to \tau(v))}))].$$

Once the choice of $F(x, y, z)$ is made, the left-hand conjunct of the right-hand side of the biconditional, namely

$$\overline{F(x, y, z)} \text{ is a formula of the form } H(x, y, z),$$

will be true, hence, *ex hypothesi*, provable in $\mathcal{S}$. Therefore $\mathcal{S}$ will prove

$$\forall x \forall y \forall z(F(x, y, z) \leftrightarrow ([(x = 0 \wedge z = y) \vee$$

$$\exists w(Nw \wedge x = sw \wedge z = \overline{\forall v(F(w, y, v) \to \tau(v))})]).$$

[17] Here I am indebted to Vann McGee (personal correspondence).

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

9.4. Some deducibilities, in $\mathcal{S}$, involving the ternary relation $F(x, y, z)$. From the last displayed sentence we can infer that the following interdeducibilities hold within the system $\mathcal{S}$:

(i) $F(0, u, a) \ \dashv\vdash_{\mathcal{S}} \ a = u$;

(ii) $F(t, u, a) \ \dashv\vdash_{\mathcal{S}}$
$(t = 0 \ \wedge \ a = u) \ \vee \ \exists w(Nw \ \wedge \ t = sw \ \wedge \ a = \overline{\forall v(F\underline{wu}v \rightarrow P(v))}\ )$.

Let us concentrate on (ii), first in the right-to-left direction:

(ii.a) $\mathcal{S}$, $(t = 0 \ \wedge \ a = u) \ \vee \ \exists w(Nw \ \wedge \ t = sw \ \wedge \ a = \overline{\forall v(F\underline{wu}v \rightarrow P(v))}\ )$
$\vdash \ F(t, u, a)$.

By the properties of disjunction, we have

(iii) $\mathcal{S}$, $\exists w(Nw \ \wedge \ t = sw \ \wedge \ a = \overline{\forall v(F\underline{wu}v \rightarrow P(v))}\ ) \ \vdash \ F(t, u, a)$.

We shall make use of a substitution instance of (iii) presently.

Now consider (ii) in the left-to-right direction, in the case where t is of the form sk. The left-hand disjunct in this case becomes $(sk = 0 \wedge a = u)$. But $\mathsf{Q}, sk = 0 \vdash \bot$; so the left-hand disjunct leads to absurdity. From $F(sk, u, a)$ it follows, then, that

$$\exists w(Nw \ \wedge \ sk = sw \ \wedge \ a = \overline{\forall v(F\underline{wu}v \rightarrow P(v))}\ ).$$

Note that $\mathsf{Q}, sk = sw \vdash k = w$. So it follows from the last claim that

$$a = \overline{\forall v(F\underline{ku}v \rightarrow P(v))}.$$

We have therefore shown that

(iv) $\mathcal{S}$, $F(sk, u, a) \ \vdash \ a = \overline{\forall v(F\underline{ku}v \rightarrow P(v))}$.

9.5. The fixed point σ. With the formula $F(x, y, z)$ in hand, McGee appeals to Theorem 3 once more, this time to find a sentence σ that is a fixed point for the unary formula

$$\neg\forall x(Nx \rightarrow \forall z(F(x, y, z) \rightarrow P(z)))$$

(with free variable y). This yields the interdeducibility

(v) $\sigma \ \dashv\vdash_{\mathcal{S}} \ \neg\forall x(Nx \rightarrow \forall z(F(x, \overline{\sigma}, z) \rightarrow P(z)))$.

Using in (iii) the substitutions $\begin{matrix} t & u & & a \\ \downarrow & \downarrow & & \downarrow \\ sb & \overline{\sigma} & & \overline{\forall v(F\underline{b\overline{\sigma}}v \rightarrow P(v))} \end{matrix}$ we obtain

OBSERVATION 2.

$$\mathcal{S} \ , \ \exists w(Nw \ \wedge \ sb = sw \ \wedge \ \overline{\forall v(F\underline{b\overline{\sigma}}v \rightarrow P(v))} = \overline{\forall v(F\underline{w\overline{\sigma}}v \rightarrow P(v))}\ )$$

$$\vdash \ F(sb, \overline{\sigma}, \overline{\forall v(F\underline{b\overline{\sigma}}v \rightarrow P(v))}\ ).$$

9.6. Back to proving lemmas.

LEMMA 13. $\mathcal{S}$, $Nb \ \vdash \ F(sb, \overline{\sigma}, \overline{\forall v(F\underline{b\overline{\sigma}}v \rightarrow P(v))}\ )$.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Proof.

$$\dfrac{\underline{\underline{Nb}}}{\dfrac{\exists!\, a}{\exists!\, \underline{\underline{sb}}}}$$

$$\dfrac{Nb \qquad sb = sb}{Nb \ \wedge \ sb = sb} \qquad \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))} = \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))}$$

$$\dfrac{Nb \ \wedge \ sb = sb \ \wedge \ \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))} = \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))}}{\underbrace{\mathcal{S} \quad , \quad \exists w(Nw \ \wedge \ sb = sw \ \wedge \ \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))} = \overline{\forall v(F\underline{w}\overline{\sigma}v \to P(v))})}_{\vdots \text{ by Observation 2}}}$$

$$F(sb, \overline{\sigma}, \overline{\forall v(F\underline{b}\overline{\sigma}v \to P(v))}\ ).$$

□

Lemma 14. $\mathcal{S} \vdash \forall x(Nx \to F(sx, \overline{\sigma}, \overline{\forall v(F(\underline{x}, \overline{\sigma}, v) \to P(v))}))$.

Proof. Immediate from Lemma 13, in which b is parametric:

$$\dfrac{\dfrac{F(sb, \overline{\sigma}, \overline{\forall v(F(\underline{b}, \overline{\sigma}, v) \to P(v))})}{Nb \to F(sb, \overline{\sigma}, \overline{\forall v(F(\underline{b}, \overline{\sigma}, v) \to P(v))})}{}^{(1)}}{\forall x(Nx \to F(sx, \overline{\sigma}, \overline{\forall v(F(\underline{x}, \overline{\sigma}, v) \to P(v))})).}$$

$$\underbrace{\mathcal{S} \quad , \quad \overline{\underline{\underline{Nb}}}^{(1)}}_{\vdots \text{ by (L13)}}$$

□

Lemma 15. $$\dfrac{\forall x(Nx \to \forall z(F(x, \overline{\sigma}, z) \to P(z)))}{\forall x(Nx \to \forall z(F(sx, \overline{\sigma}, z) \to P(z))).}$$

Proof. In Lemma 11, use the substitution

$$\begin{array}{c} \Psi x \\ \downarrow \\ \forall z(F(x, \overline{\sigma}, z) \to P(z)). \end{array}$$

□

Lemma 16.

$$\dfrac{\forall x(Nx \to \forall z(F(sx, \overline{\sigma}, z) \to P(z))) \ \ \forall x(Nx \to F(sx, \overline{\sigma}, \overline{\forall z(F(\underline{x}, \overline{\sigma}, z) \to P(z))}))}{\forall x(Nx \to P[\forall z(F(\underline{x}, \overline{\sigma}, z) \to P(z))]).}$$

Proof. In Lemma 12:

$$\dfrac{\forall x(Nx \to \forall z(\Xi xz \to \Theta z)) \quad \forall x(Nx \to \Xi xt)}{\forall x(Nx \to \Theta t)}$$

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

use the substitutions

$$\begin{array}{ccc} \Xi xz & \Theta & t \\ \downarrow & \downarrow & \downarrow \\ F(x,\overline{\sigma},z) & P & \overline{\forall z(F(\underline{x},\overline{\sigma},z)\to P(z))}. \end{array}$$

□

LEMMA 17.

$$\frac{\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))\quad \forall x(Nx\to F(sx,\overline{\sigma},\overline{\forall z(F(\underline{x},\overline{\sigma},z)\to P(z))}))}{P[\forall x(Nx\to\forall z(F(\underline{x},\overline{\sigma},z)\to P(z)))].}$$

Proof. Immediate by (4) and Lemmas 15 and 16:

$$\underbrace{\mathcal{S}\ ,\ \dfrac{\dfrac{\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}{\forall x(Nx\to\forall z(F(sx,\overline{\sigma},z)\to P(z)))}(\text{L15})\qquad \forall x(Nx\to F(sx,\overline{\sigma},\overline{\forall z(F(\underline{x},\overline{\sigma},z)\to P(z))}))}{\forall x(Nx\to P[\forall z(F(\underline{x},\overline{\sigma},z)\to P(z))])}(\text{L16})}_{\vdots\ \text{by (4)}}$$

$$P[\forall x(Nx\to\forall z(F(\underline{x},\overline{\sigma},z)\to P(z)))].$$

□

LEMMA 18. $\mathcal{S},\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))\ \vdash\ P[\sigma]$.

Proof.

$$\dfrac{\underbrace{\mathcal{S}\ ,\ \overline{\overline{\sigma}=\overline{\sigma}}}_{\vdots\ \text{by (i)}}\qquad \dfrac{\dfrac{\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}{N0\to\forall z(F(0,\overline{\sigma},z)\to P(z))}\qquad \overline{N0}}{\dfrac{\forall z(F(0,\overline{\sigma},z)\to P(z))}{F(0,\overline{\sigma},\overline{\sigma})\to P[\overline{\sigma}]}}}{F(0,\overline{\sigma},\overline{\sigma})\qquad F(0,\overline{\sigma},\overline{\sigma})\to P[\overline{\sigma}]}$$

$$P[\sigma].$$

□

LEMMA 19. $\mathcal{S}\ \vdash\ P[\sigma\to\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))]$.

Proof.

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\ ,\ \overset{(1)}{\overline{\sigma}}}_{\vdots\ \text{by (v)}}}{\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}(1)}{\sigma\to\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}(1)$$

$$P[\sigma\to\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))].$$

□

LEMMA 20. $\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))\ \vdash$
$P[\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))]$.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

Proof. Immediate from (2) and Lemmas 18 and 19:

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}}{\vdots\ (\text{L18})} \qquad \dfrac{\mathcal{S}}{\vdots\ (\text{L19})} \qquad \dfrac{P[\sigma] \qquad P[\sigma\to\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))]}{P[\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))].}(2)$$

□

LEMMA 21. $\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))\ \vdash$
$P[\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))]$.

Proof. Immediate by Lemmas 14, 15, and 17:

$$\dfrac{\underbrace{\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}\ ,\ \dfrac{\dfrac{\mathcal{S}}{\vdots\ (\text{L14})}}{\dfrac{\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}{\forall x(Nx\to\forall z(F(sx,\overline{\sigma},z)\to P(z)))}(\text{L15})}}{\vdots\ (\text{L17})}$$
$$P[\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))].$$

□

LEMMA 22. $\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))\ \vdash\ \bot$.

Proof. Immediate by (3) and Lemmas 20 and 21:

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}}{\vdots\ (\text{L20})} \qquad \dfrac{\underbrace{\mathcal{S}\ ,\ \forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}}{\vdots\ (\text{L21})} \qquad \dfrac{P[\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))] \quad P[\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))]}{\bot}(3).$$

□

LEMMA 23. $\mathcal{S}\vdash\sigma$.

Proof. Immediate by (v) and Lemma 22:

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\quad ,\quad \dfrac{}{\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}(1)}}{\vdots\ (\text{L22})} \qquad \dfrac{\dfrac{\bot}{\underbrace{\neg\forall x(Nx\to\forall z(F(x,\overline{\sigma},z)\to P(z)))}}(1)}{\vdots\ (\text{v})}$$
$$\sigma.$$

□

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

LEMMA 24. $\mathcal{S} \vdash P[\sigma]$, *i.e.*, $\mathcal{S} \vdash P(\overline{\sigma})$.

Proof. Immediate from Lemma 23 by Condition 1. □

LEMMA 25. $\mathcal{S} \vdash \forall z(F(0, \overline{\sigma}, z) \to P(z))$.

Proof.

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\ ,\ \overline{F(0,\overline{\sigma},a)}^{(1)}}_{\vdots\ \text{by (i)}} \qquad \underset{\vdots\ \text{(L24)}}{\mathcal{S}}}{\dfrac{a = \overline{\sigma} \qquad P(\overline{\sigma})}{P(a)}}}{\dfrac{F(0,\overline{\sigma},a) \to P(a)}{\forall z(F(0,\overline{\sigma},z) \to P(z)).}}\ (1)$$

□

LEMMA 26. $\mathcal{S} \vdash P[\forall z(F(0, \overline{\sigma}, z) \to P(z))]$, *i.e.*, $\mathcal{S} \vdash P(\overline{\forall z(F(0, \overline{\sigma}, z) \to P(z))})$.

Proof. Immediate from Lemma 25 by Condition 1. □

LEMMA 27. *For all n, we have* $\mathcal{S} \vdash \forall z(F(\underline{n}, \overline{\sigma}, z) \to P(z))$.

Proof. By mathematical induction on n. The basis is accomplished by Lemma 25. Now assume the *Inductive Hypothesis* that

$$\mathcal{S} \vdash \forall z(F(\underline{k}, \overline{\sigma}, z) \to P(z)).$$

It follows by Condition 1 that

(vi) $\mathcal{S} \vdash P[\overline{\forall z(F(\underline{k}, \overline{\sigma}, z) \to P(z))}]$.

For the inductive step, consider the following proof within the system $\mathcal{S}$:

$$\dfrac{\dfrac{\underbrace{\mathcal{S}\ \ ,\ \ \overline{F(\underline{sk},\overline{\sigma},a)}^{(1)}}_{\vdots\ \text{by (iv)}} \qquad \underset{\vdots\ \text{by (vi)}}{\mathcal{S}}}{\dfrac{a = \overline{\sigma} \qquad P(\overline{\sigma})}{P(a)}}}{\dfrac{F(\underline{sk},\overline{\sigma},a) \to P(a)}{\forall z(F(\underline{sk},\overline{\sigma},z) \to P(z)).}}\ (1)$$

Thus $\mathcal{S} \vdash \forall z(F(\underline{sk}, \overline{\sigma}, z) \to P(z))$. The result now follows by induction. □

Recall Lemma 22: $\mathcal{S}\ ,\ \forall x(Nx \to \forall z(F(x, \overline{\sigma}, z) \to P(z))) \vdash \bot$. It follows that $\mathcal{S}$ is ω-inconsistent. This completes our constructive proof of Theorem 8.

§10. Concluding summary. We gather here the main results constructively proved in full formal detail above. This is for the convenience of the reader who might wish, on the other side of the topical divide, to pursue 'paradoxicality analyses' of any of them.

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

We have seen that in the case of Theorem 4 both sides of the divide have been worked out. We remind the reader that in our exposition of Theorems 5–8 we have been working on the metamathematical and arithmetical side of the topical divide, establishing the straightforward impossibility of furnishing definitions, within the language of arithmetic, of predicates P satisfying the respective and various combinations of conditions. Below is our summary list of those theorems, with their respective sets of P-conditions. We remind the reader that in these metamathematical theorems there is no occasion to discuss matters of normal form or normalizability. Those matters come into play only on the other side of the topical divide, when one studies the structures of the disproofs that are involved as regimentations of the reasoning involved in diagnosing logico-semantic paradoxicality.

THEOREM 4 (Tarski). *Suppose that the formal system $\mathcal{S}$ is sufficiently strong, and that it has a predicate P that satisfies the rules of P-Introduction and P-Elimination*:

$$\mathcal{S}, \theta \vdash P[\theta] \quad \textit{and} \quad \mathcal{S}, P[\theta] \vdash \theta.$$

Then $\mathcal{S}$ is inconsistent.

THEOREM 5 (Löb). *For all φ, if $\mathcal{S}, P\varphi \vdash \varphi$ then $\mathcal{S} \vdash \varphi$.*

THEOREM 6 (Montague). *Any theory $\mathcal{S}$ that interprets* Q *and whose provability predicate P satisfies the following four conditions is inconsistent*:

1. $\mathcal{S} \vdash P[P[\theta] \rightarrow \theta]$.
2. $\dfrac{\vdash \theta}{\mathcal{S} \vdash P[\theta].}$
3. $\dfrac{\mathcal{S} \vdash P[\varphi \rightarrow \psi] \quad \mathcal{S} \vdash P[\varphi]}{\mathcal{S} \vdash P[\psi].}$
4. $\mathcal{S}, P[\theta] \vdash \theta$.

THEOREM 7 (McGee). *Suppose $\mathcal{S}$ is sufficiently strong, and that its provability predicate P satisfies the introductory and eliminative conditions*

$$\frac{\mathcal{S} \vdash \theta}{\mathcal{S} \vdash P[\theta]} \quad \textit{and} \quad \mathcal{S}, P[\theta] \vdash \theta.$$

Then $\mathcal{S}$ is inconsistent.

THEOREM 8 (McGee). *Suppose that $\mathcal{S}$ interprets* Q *and satisfies the four conditions stated below. Then $\mathcal{S}$ is ω-inconsistent.*

1. $\dfrac{\mathcal{S} \vdash \varphi}{\mathcal{S} \vdash P\varphi.}$
2. $\mathcal{S}, P\varphi, P[\varphi \rightarrow \psi] \vdash P\psi$.
3. $\mathcal{S}, P[\neg\varphi], P[\varphi] \vdash \bot$.
4. $\mathcal{S}, \forall x(Nx \rightarrow P[\varphi\underline{x}]) \vdash P[\forall x(Nx \rightarrow \varphi x)]$.

On the other side of the topical divide, we conjecture, would be complementing revelations of the logico-semantic paradoxicality that would ensue if one were to take these various sets of conditions as new rules of inference directly governing a freshly adopted primitive predicate P (in an extension of the language of arithmetic). That would require demonstrating that the new rules of inference would engender

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press

paradoxical disproofs, where the paradoxicality consists in their *non-normalizability*. The reduction sequences emanating from such a disproof would fail to terminate.

We commend this unfinished business as a challenging and potentially rewarding research program.

Acknowledgments. The author is grateful to two anonymous referees for constructive comments that have improved the presentation. All errors or deficiencies that remain are the sole responsibility of the author.

BIBLIOGRAPHY

[1] Buss, S. R. (1998). First-order proof theory of arithmetic. In Buss, S. R., editor. *Handbook of Proof Theory*. Amsterdam: Elsevier Science B. V, pp. 79–147.

[2] Friedman, H. (1978). Classically and intuitionistically provably recursive functions. In Müller, G. H., and Scott, D. S., editors. *Higher Set Theory: Proceedings, Oberwolfach, Germany, April 13–23*. Springer, Berlin, Heidelberg, pp. 21–27.

[3] Löb, M. H. (1955). Solution of a problem of Leon Henkin. *Journal of Symbolic Logic*, **20**(2), 115–118. https://doi.org/10.2307/2266895.

[4] McGee, V. (1991). *Truth, Vagueness and Paradox: An Essay on the Logic of Truth*. Indianapolis: Hackett Publishing Co.

[5] Montague, R. (1974). Syntactical treatments of modality, with corollaries on reflection principles and finite axiomatizability (1963). In Thomason, R., editor. *Formal Philosophy: Selected Papers of Richard Montague*. New Haven: Yale University Press, pp. 286–302.

[6] Tarski, A., Mostowski, A., & Robinson, R. M. (1953). *Undecidable Theories*. Amsterdam: North Holland.

[7] Tennant, N. (1982). Proof and paradox. *Dialectica*, **36**, 265–296. https://doi.org/10.1111/j.1746-8361.1982.tb00820.x.

[8] ———. (1995). On paradox without self-reference. *Analysis*, **55**(3), 199–207.

[9] ———. (2015). A new unified account of truth and paradox. *Mind*, **123**, 571–605.

[10] ———. (2017). *Core Logic*. Oxford: Oxford University Press.

[11] ———. (2023). Core Gödel. *Notre Dame Journal of Formal Logic*, **64**(1), 15–59. https://doi.org/10.1215/00294527-2022-0033.

[12] ———. (2024). Which 'intensional paradoxes' are paradoxes? *Journal of Philosophical Logic*, **53**, 933–957. https://doi.org/10.1007/s10992-024-09761-8.

[13] ———. (2024). Core Tarski and core McGee. *Notre Dame Journal of Formal Logic*, **66**(1), 31–55. https://doi.org/10.1215/00294527-2024-0032.

THE OHIO STATE UNIVERSITY
UNITED STATES
E-mail: tennant.9@osu.edu

https://doi.org/10.1017/S175502032510097X Published online by Cambridge University Press