BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 6, Number 3, May 1982

FIXED POINT ALGEBRAS

BY C. SMORYŃSKI

Although self-reference in arithmetic was used to impressive effect by Gödel in 1930 (published in 1931) when he noted the sentence asserting its own unprovability to be unprovable, and although this use immediately appealed to philosophers and philosophical logicians, it has largely been ignored by mathematical logicians. Indeed, it is only in the 1970s that arithmetic self-reference has begun to be systematically studied and applied. One aspect of this study is algebraic.

In simplest terms one can distinguish two types of self-reference—extensional and nonextensional. Extensional self-reference lends itself quite readily to algebraic description and modelling, with some types of extensional self-reference even being amenable to algebraic study. The purpose of the present paper is to expound upon this algebraic modelling, touching briefly on its successes and delineating roughly the limits to this success. The central notion of this exposition is that of a fixed point algebra. This notion is a new one—it is untested and, hence, of only provisional interest. But it does appear useful: It makes the present discussion cohere reasonably well; it provides a convenient framework in which to find and formulate questions about extensional self-reference; and it has allowed me to prove a theorem on the limitations of the use of finite algebras in studying such self-reference. Moreover, the algebra is fairly pleasant and provides a philosophically neutral framework in which to discuss arithmetic self-reference.

While the basic context from which the notion of a fixed point algebra arises is logical, the concept itself is algebraic and the following treatment is almost entirely algebraic. I have included logical material, including references to some rather arcane results of mathematical logic, in the discussion; but I have tried to keep this at a minimum. As a consequence, I declare most of the present paper accessible to the general mathematician who possesses only a small knowledge of boolean algebras. Only the logical asides (on background, motivation, and occasional applications) and the section on infinite fixed point algebras (for which I possess no nontrivial nonlogical examples) should be meaningless to the nonspecialist. [On the other hand, I must admit that, as I introduced fixed point algebras to serve as a vehicle for briefly surveying some of the algebraic aspects of self-reference in arithmetic, there isn't much left to the paper when the logic has been excised.]

Received by the editors September 15, 1981.
1980 *Mathematics Subject Classification.* Primary 03-02, 03G05; Secondary 03B45, 03F30.

0273-0979/81/0000-0275/$11.00

In the immediately following section I discuss the logical background to this paper and present a formal definition of fixed point algebras. The ensuing sections (§§2–5) devote themselves mainly to the study of finite fixed point algebras. This study begins with a review of diagonalisable algebras and their use in giving a successful analysis of arithmetic provability and ends with the Fundamental Theorem of Finite Fixed Point Algebras, which asserts vaguely that one cannot do much better. In the final section I make a few remarks about infinite fixed point algebras.

I have benefited both in the research behind and the exposition of the present paper by several discussions with D. H. J. de Jongh and D. Myers.

1. Fixed point algebras. One of the central concerns of logic is the use of language in mathematics. Now, languages are given by alphabets, words constructed over the alphabets, formulae or sentences constructed from these words, etc. In short, languages are largely matters of finite strings of symbols and these strings can be put in one-one correspondence with the natural numbers—to each formula φ of a given language there corresponds a natural number $\ulcorner\varphi\urcorner$. It happens that this correspondence can usually be accomplished in such a way that natural syntactic properties of φ, ψ, etc., transform to natural arithmetic properties of $\ulcorner\varphi\urcorner$, $\ulcorner\psi\urcorner$, etc. Thus, as Gödel observed, if the language contains arithmetic, it is capable of discussing itself.

As mathematics concerns itself with proofs, mathematical logic concerns itself with formal representations of proofs in formal languages. Formal theories in languages containing arithmetic which are strong enough to prove some basic arithmetic propositions will be strong enough to discuss and provably discuss their own syntax. This is what Gödel [**9**] observed: If T is a formal theory containing a sufficient amount of arithmetic (e.g. the logicians' favourites: PA (Peano arithmetic), ZF (Zermelo-Fraenkel set theory), GB (Gödel-Bernays class theory), and KM (Kelley-Morse class theory)), there is a representation $\mathrm{Pr}_T(\cdot)$ of provability in T, i.e. a formula $\mathrm{Pr}_T(v)$ with only the variable v free such that, for all sentences φ of the language of T,

$$T \vdash \varphi \Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner\varphi\urcorner), \tag{1}$$

where the turnstile, $\vdash$, denotes provability.

One further observation yields incompleteness. A basic syntactic operation is the substitution of terms for variables. This means that the function taking one from $\ulcorner\varphi v\urcorner$ and $\ulcorner t\urcorner$ to $\ulcorner\varphi t\urcorner$ can be simulated in the theory T. In particular, one can toy with the diagonal function $\ulcorner\varphi v\urcorner \mapsto \ulcorner\varphi(\ulcorner\varphi v\urcorner)\urcorner$ and obtain the

1.1. Diagonalisation Lemma. *Let ψv have only v free. There is a sentence φ such that $T \vdash \varphi \leftrightarrow \psi(\ulcorner\varphi\urcorner)$.*

Proof. Let Δ be the diagonal function and, for $\theta v = \psi(\Delta v)$, let $n = \ulcorner\theta v\urcorner$ and $\varphi = \theta n$. Then

$$T \vdash \varphi \leftrightarrow \theta(\ulcorner\theta v\urcorner) \leftrightarrow \psi(\Delta(\ulcorner\theta v\urcorner)) \leftrightarrow \psi(\ulcorner\theta(\ulcorner\theta v\urcorner)\urcorner) \leftrightarrow \psi(\ulcorner\theta n\urcorner) \leftrightarrow \psi(\ulcorner\varphi\urcorner).$$

Q.E.D.

1.2. COROLLARY (FIRST INCOMPLETENESS THEOREM). *If T is consistent, there is a true sentence φ which is not provable in T. In particular, if T is a true theory, $\neg\varphi$ is also not provable; i.e. φ is formally undecidable in T.*

PROOF. Choose φ as in the proof of the Diagonalisation Lemma so that $T \vdash \varphi \leftrightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$. Since φ asserts its own unprovability, it suffices to show φ to be unprovable. To see this, note

$$\begin{aligned} T \vdash \varphi &\Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner), \quad \text{by (1)}, \\ &\Rightarrow T \vdash \neg\varphi, \quad \text{by definition of } \varphi, \\ &\Rightarrow T \text{ is inconsistent}, \end{aligned}$$

contrary to assumption. Q.E.D.

Assuming the representation Pr_T of provability in T to be especially nice, one gets two further interesting corollaries.

1.3. COROLLARY (SECOND INCOMPLETENESS THEOREM). *If T is consistent, then $T \nvdash \mathrm{Con}_T$, where Con_T is the assertion, $\neg \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner)$, that T is consistent.*

1.4. COROLLARY (FIXED POINT CALCULATION). *Let $T \vdash \varphi \leftrightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$. Then $T \vdash \varphi \leftrightarrow \mathrm{Con}_T$.*

The particular niceties necessary beyond assumption (1) above are the assumptions (2) and (3).

$$(2) \qquad T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \wedge \mathrm{Pr}_T(\ulcorner \varphi \to \psi \urcorner) \to \mathrm{Pr}_T(\ulcorner \psi \urcorner),$$

$$(3) \qquad T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T\big(\ulcorner \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner\big).$$

[Condition (2) is trivially built into a representation; (3), the formalisation of (1), requires some strength to establish.]

PROOF OF 1.3. It suffices to show $T \vdash \mathrm{Con}_T \to \varphi$ for any sentence φ such that $T \vdash \varphi \leftrightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$. To this end, note that a few applications of (1) and (2) yield

$$(4) \qquad T \vdash \neg\varphi \leftrightarrow \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner \neg\varphi \urcorner) \leftrightarrow \mathrm{Pr}_T\big(\ulcorner \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner\big)$$

and

$$(5) \quad T \vdash \varphi \wedge \neg\varphi \to 0 = 1 \Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \wedge \mathrm{Pr}_T(\ulcorner \neg\varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner).$$

But, by (3),

$$T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T\big(\ulcorner \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner\big),$$

which with (4) yields

$$T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner \neg\varphi \urcorner),$$

whence $T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \wedge \mathrm{Pr}_T(\ulcorner \neg\varphi \urcorner)$. But then (5) yields

$$T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner),$$

the contraposition of which yields the Corollary. Q.E.D.

PROOF OF 1.4. Again (1) and (2) yield

$$T \vdash 0 = 1 \to \varphi \Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner) \to \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$$

for any sentence φ. Thus, $T \vdash \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Con}_T$. If one now assumes $T \vdash \varphi \leftrightarrow \neg \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$, the proof of 1.3 yields the Corollary. Q.E.D.

The thing to notice about these proofs is that, modulo the derivations of the Diagonalisation Lemma and the Derivability Conditions (1)–(3), they are fairly algebraic. I don't mean this in a merely heuristic sense—a few simple definitions transform these questions and proofs into purely algebraic matters.

1.5. DEFINITION. Let T be a formal theory. The *Lindenbaum algebra* of T is the algebra A_T of equivalence classes of sentences (i.e. formulae with no free variables) under the relation of provable equivalence. More precisely, the elements of A_T are the classes

$$[\varphi] = \{\text{sentences } \psi \text{ in the language of } T\colon T \vdash \varphi \leftrightarrow \psi\}$$

and the boolean operations are the induced ones, e.g.

$$[\varphi] \wedge [\psi] = [\varphi \wedge \psi], \qquad 0 = [\varphi \wedge \neg\varphi], \qquad 1 = [\varphi \vee \neg\varphi].$$

(Note that T is consistent iff $0 \neq 1$ in A_T.)

Using the first two derivability conditions, (1) and (2), we see that $\mathrm{Pr}_T(\cdot)$ preserves equivalence, i.e.

$$T \vdash \varphi \leftrightarrow \psi \Rightarrow T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \leftrightarrow \mathrm{Pr}_T(\ulcorner \psi \urcorner),$$

and so induces an operation $\tau\colon A_T \to A_T$ defined by

$$\tau[\varphi] = [\mathrm{Pr}_T(\ulcorner \varphi \urcorner)].$$

The derivability conditions themselves assume an algebraic flavour:

$$(1') \qquad \tau 1 = 1,$$

$$(2') \qquad \tau x \wedge \tau(x \to y) \leqslant \tau y,$$

$$(3') \qquad \tau x \leqslant \tau\tau x.$$

Finally, the Diagonalisation Lemma becomes an assertion about the existence of fixed points to special polynomials in τ:

1.6. LEMMA. *Every polynomial $\alpha(x)$ over A_T in τ in which the variable x occurs only within the scope of a τ possesses a fixed point $a \in A_T$: $\alpha a = a$.*

The notion of polynomial used here is simply that from universal algebra: Polynomials are functions arising from constant functions and the identity function (represented by the variable x) by means of the boolean operations and τ. Replacing the identity function as a generator by all polynomials of the form $\tau(\beta x)$, where β is a polynomial, one obtains from the above the subclass of all polynomials mentioned in the Lemma.

The validity of Lemma 1.6 follows from that of the Diagonalisation Lemma because occurrences of x inside the scopes of τ's correspond to occurrences of φ in contexts of the form $\mathrm{Pr}_T(\ulcorner \cdots \varphi \cdots \urcorner)$, i.e. of the form $\mathrm{Pr}_T(f(\ulcorner \varphi \urcorner))$. One simply pulls the equation $x = \alpha x$ back to an equivalence $\varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$.

Finally, I note the algebraic formulations of Corollaries 1.2–1.4.

1.2′. COROLLARY. *Let* $a = \neg\tau a$.
(i) *If* $0 \neq 1$ *in* A_T, *then* $a \neq 1$.
(ii) *If, for all* $b \in A_T$, $\tau b = 1$ *implies* $b = 1$, *then* $\neg a \neq 1$.

PROOF. (i). Assume $a = 1$. By (1′), $\tau a = 1$. But then $a = \neg\tau a = \neg 1 = 0$, a contradiction.
(ii) Note

$$\neg a = 1 \Rightarrow \tau a = 1 \Rightarrow a = 1,$$

again a contradiction. Q.E.D.

1.3′. COROLLARY. *If* $0 \neq 1$ *in* A_T, $\tau 0 \neq 0$.

PROOF. Let $a = \neg\tau a$. Then $\neg a = \tau a$ and $\tau(\neg a) = \tau\tau a$. But $\neg a = a \to 0$ and $\tau a \leqslant \tau\tau a = \tau(\neg a) = \tau(a \to 0)$, whence $\tau a = \tau a \wedge \tau(a \to 0)$. But $\tau a \wedge \tau(a \to 0) \leqslant \tau 0$, i.e. $\tau a \leqslant \tau 0$. If $\tau 0 = 0$, then $a = \neg\tau a = 1$, which fails by 1.2′. Q.E.D.

1.4′. COROLLARY. *In* A_T, *if* $a = \neg\tau a$, *then* $a = \neg\tau 0$.

The reader can supply the details. He might also wish to try his hand with the following result from [**17**].

1.7. EXERCISE (LÖB'S THEOREM). Show that, for all $a \in A_T$, $\tau(\tau a \to a) \leqslant \tau a$.

[HINT. Let $b = \tau(b \to a)$. Alternatively, use $b = \tau b \to a$ to show: If $\tau a \leqslant a$, then $a = 1$.]

I do not expect the reader to applaud this as a great algebraic success. Indeed, it is not very impressive. But it does raise a few questions—most significantly: How far can such simple elementary methods go? Put differently: Is there a genuine algebraic aspect to arithmetic self-reference? What types of self-reference fall into the algebraic sphere? And, what can or cannot an algebraic analysis tell us about self-reference? These are the types of questions I would like to address here. At the moment, however, the most fruitful question to ask is: What kind of algebra arises from the study of self-reference? Or even: What kinds of algebras arise from this study?

One such type of algebra is a diagonalised algebra:

1.8. DEFINITION. A structure $(A; \tau)$, where A is a boolean algebra (whose boolean operations we have notationally suppressed), is a *diagonalised algebra* if it satisfies

(i) $\tau 1 = 1$,
(ii) for all $x, y \in A$, $\tau x \wedge \tau(x \to y) \leqslant \tau y$,
(iii) for all $x \in A$, $\tau x \leqslant \tau\tau x$, and
(iv) every polynomial $\alpha(x)$ in which the variable x occurs only inside the scopes of τ's possesses a fixed point $a \in A$: $\alpha a = a$.

Diagonalised algebras, introduced by Magari and his followers, largely capture the essence of Pr_T. It turns out that the diagonalised algebras form an equational variety for which both the class of finite algebras and the Lindenbaum algebra $(A_{\mathrm{PA}}; \tau)$ are equationally generic. Thus, arithmetic provability is amenable to study by finite algebras.

The success doesn't end here. Other structures equationally generic for the variety of diagonalised algebras are $(A_{ZF}; \tau)$, $(A_{GB}; \tau)$, and $(A_{ZF}; \sigma)$, where τ denotes the respective provability operators of A_{ZF} and A_{GB}, and σ the operator

$$\sigma[\varphi] = [\text{“}\varphi \text{ is true in all } \omega\text{-models of ZF”}].$$

Moreover, other structures are known equationally generic for special subvarieties of the variety of diagonalised algebras. Any equationally expressible question about these structures is directly algebraically decidable.

However, the list of known examples of τ-operators on Lindenbaum algebras is quite small. A systematic enumeration has yet to be started. Moreover, the question of what varieties emerge when such operators are taken two or three at a time has not even been addressed. For example, aside from copying the τ-axioms for each σ and τ in $(A_{ZF}; \sigma, \tau)$, with σ, τ as above, and strengthening the diagonalisation to allow for fixed points to polynomials in which the occurrences of x are split among the scopes of σ and τ, what new equations must be added? The only relation between σ and τ that comes to mind is: $\tau x \leqslant \sigma x$. Is this sufficient?

In addition to accomplishing all this, one must analyse operators other than τ-operators. A good candidate for analysis is the operator ρ on A_{GB} given by

$$\rho[\varphi] = [\text{“GB} + \varphi \text{ is relatively interpretable in GB”}].$$

As with any finitely axiomatised theory, over GB relative interpretability is something like an infinite disjunction of provability assertions. Can the study of $(A_{GB}; \rho)$ be reduced to the study of finite "multiply diagonalised" algebras $(A; \tau_1, \ldots, \tau_n)$ by interpreting ρ as a disjunction of some of the τ-operators?

Finally, there are algebraically meaningful operators on Lindenbaum algebras which are demonstrably not amenable to study by finite algebras. It follows for PA, ZF and KM by a result of Orey (cf. [**6**] and [**32**]) and for GB by a result of Solovay [**51**] that there are homomorphisms of the Lindenbaum algebras definable in the given languages. As we shall see, the corresponding algebras, say $(A_T; \mu)$, are not adequately modelled by any finite algebras.

The positive approach to the question of delineating the extent of possible algebraic success in studying self-reference, in establishing lower bounds to the limit of this success as it were, promises to be rather pedestrian in pace: First one studies this; then one studies that;... And the list is not short. In any event, this success has not yet got very far and there is little, at this point in time, to report on about these lower bounds. The opposite question of upper bounds brings us back to the question: What algebras arise from the study of self-reference?

1.9. DEFINITIONS. Let T be a formal theory. A formula χv of the language of T which possesses only the variable v free is *extensional* if, for all sentences φ, ψ,

$$T \vdash \varphi \leftrightarrow \psi \Rightarrow T \vdash \chi(\ulcorner \varphi \urcorner) \leftrightarrow \chi(\ulcorner \psi \urcorner).$$

Two extensional formulae $\chi_1 v$, $\chi_2 v$ are *equivalent* if, for each sentence φ, $T \vdash \chi_1(\ulcorner \varphi \urcorner) \leftrightarrow \chi_2(\ulcorner \varphi \urcorner)$. The equivalence class of an extensional formula χv is denoted $[\chi v]$. The *Lindenbaum formula algebra* of T, B_T, is the collection of these equivalence classes under the induced boolean operations. The *Lindenbaum fixed point algebra* of T is the structure (B_T, A_T), with the obvious boolean operations and with the application operation: for $[\chi v] \in B_T$ and $[\varphi] \in A_T$.

$$[\chi v]([\varphi]) = [\chi(\ulcorner \varphi \urcorner)].$$

[A TINY REMARK. A logically more natural equivalence between formulae $\chi_1 v$ and $\chi_2 v$ is given by simple provable equivalence: $T \vdash \forall v\ [\chi_1 v \leftrightarrow \chi_2 v]$. Algebraically, however, the weaker pointwise equivalence is more natural. For one thing, it yields $[\chi_1 v] = [\chi_2 v]$ iff the functions induced by these classes under the above definition are identical.]

Bearing in mind that one might wish to consider subalgebras of the full Lindenbaum fixed point algebras of various theories, the concept to be abstracted from the algebras (B_T, A_T) is that following:

1.10. DEFINITION. A *fixed point algebra* (FPA) is a pair (B, A) of boolean algebras such that

(i) each $\alpha \in B$ is a function $\alpha: A \to A$,

(ii) B contains the constant functions $\lambda x.a$ for each $a \in A$,

(iii) the boolean operations on B are pointwise on A,

(iv) B is closed under composition, and

(v) each $\alpha \in B$ has a fixed point $a \in A$: $\alpha a = a$.

While diagonalised algebras are not FPA's they can easily be transformed into FPA's: Given a diagonalised algebra $(A; \tau)$, let B be the set of polynomials $\alpha(x)$ over $(A; \tau)$ in which all occurrences of x lie within the scopes of τ's. By Lemma 1.6, under the obvious choice of operations, (B, A) is an FPA. Due to the inter-reducibility of problems about the structures $(A; \tau)$ and (B, A), I declare it harmless to ambiguously call both structures diagonalised algebras and shall do so. An algebra (B_0, A), where $B_0 \subseteq B$ and (B, A) is diagonalised, will be called *subdiagonalisable*. (The change in suffix will become self-explanatory early in the next section.)

The rest of the paper is divided into two parts. The first, longer, part is devoted to finite FPA's. This begins with a review of (finite and infinite) diagonalisable algebras (the equational reformulation of diagonalised algebras), including their applications and representation theory, and continue through some lemmas on finite FPA's to the introduction of *closed* FPA's—FPA's satisfying an additional closure condition—and the proof that all finite closed FPA's are subdiagonalisable. As the additional closure condition is logically quite natural, this result puts a severe limitation on any further use of finite FPA's in studying self-reference. The final, rather anticlimactic, section constitutes the second part of the paper. It contains a few observations on infinite FPA's, mostly on Lindenbaum FPA's.

Before proceeding, I should make some conventions to kill the ambiguity among boolean and both formal logical and informal logical notations. Rather

than using $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$ in all three contexts, I shall adopt the following trinity:

informal logical $\sim$	formal logical $\neg$	boolean $(\cdot)'$
&	$\wedge$	$\cdot$
or	$\vee$	$+$
$\Rightarrow$	$\supset$	$\rightarrow$
$\Leftrightarrow$	$\equiv$	$\leftrightarrow$
$\forall$	$\forall$	
$\exists$	$\exists$	

Moreover, given an FPA (B, A), I use greek letters $\alpha, \beta, \gamma, \ldots$ for elements of B, and roman letters $a, b, c, \ldots, x, y, z$ for elements of A, with letters early in the alphabet usually denoting fixed elements and letters late in the alphabet indicating greater variability.

2. Diagonalisable algebras. As mentioned in the last section, the family of diagonalised algebras forms an equational variety.

2.1. DEFINITION. A structure $(A; \tau)$, where A is a boolean algebra is a *diagonalisable algebra* (DA) if it satisfies

(i) $\tau 1 = 1$,

(ii) $\forall x, y \in A, \tau x \cdot \tau(x \rightarrow y) \leqslant \tau y$,

(iii) $\forall x \in A, \tau x \leqslant \tau\tau x$, and

(iv) $\forall x \in A, \tau(\tau x \rightarrow x) \leqslant \tau x$.

2.2. THEOREM. *The classes of diagonalised and diagonalisable algebras coincide.*

Half of this—the diagonalisable nature of diagonalised algebras—follows immediately from Löb's Theorem (Exercise 1.7, above). The other half was proven independently by de Jongh (cf. [**42**] or [**43**]) and Sambin [**36**]. While the exposition of this proof takes us a bit off course, it is simple enough not to cause too great a digression. Moreover, it is a nice application of algebra to logic.

We first need a definition and a couple of lemmas

2.3. DEFINITION. Let $(A; \tau)$ be a DA. A subset $F \subseteq A$ is a *τ-filter* if it satisfies

(i) $0 \notin F$,

(ii) $\forall x, y \in A, x, y \in F \Rightarrow x \cdot y \in F$,

(iii) $\forall x, y \in A, x \in F \,\&\, x \leqslant y \Rightarrow y \in F$, and

(iv) $\forall x \in A, x \in F \Rightarrow \tau x \in F$.

2.4. EXAMPLES. Let $(A; \tau)$ be a *DA*, $a \in A$. Then each of the intervals

$$[\tau a, 1] = \{x \in A : \tau a \leqslant x\} \quad \text{and} \quad [a \cdot \tau a, 1] = \{x \in A : a \cdot \tau a \leqslant x\}$$

is a τ-filter.

2.5. LEMMA. *Let* $(A; \tau)$ *be a* DA *and* $F \subseteq A$ *a* τ*-filter. Define* $(A/F; \tau/F)$ *by*
(i) $x/F = \{y \in A: x \leftrightarrow y \in F\}$,
(ii) $x/F \cdot y/F = (x \cdot y)/F$,
(iii) $(x/F)' = (x')/F$,
(iv) $(\tau/F)(x/F) = (\tau x)/F$.
Then $(A/F; \tau/F)$ *is a* DA.

Since the DA's form an equational variety and equations are preserved under quotients, it suffices to show the operations of A/F to be well defined. This is routine.

With this, the proof of Theorem 2.2 becomes simply a computation, albeit an inspired one: The fact is that the Fixed Point Calculation (Corollary 1.4) of the previous section was not an isolated observation, but a prototype—every such instance of self-reference has explicitly definable fixed points. One has but to find them.

2.6. THEOREM (DE JONGH-SAMBIN THEOREM). *Let* $\alpha(x, x_1, \ldots, x_n)$ *be a* τ*-polynomial in several variables in which each occurrence of* x *lies within the scope of a* τ. *There is a polynomial* $\delta(x_1, \ldots, x_n)$ *such that, in all* DA's $(A; \tau)$,

$$\delta(x_1, \ldots, x_n) = \alpha(\delta(x_1, \ldots, x_n), x_1, \ldots, x_n).$$

PROOF. The proof is an inductive one.

Pre-basis. Let $\alpha x = \tau(\gamma x)$ for a polynomial γx, where mention of the variables $x_1, \ldots, x_n$ is supressed. Then $\delta = \alpha 1$. To see this, let $F = [\alpha 1, 1] = [\tau\gamma 1, 1]$ be the principal τ-filter generated by $\alpha 1$. Note

$$\begin{aligned} \alpha 1 \in F &\Rightarrow (\alpha 1)/F = 1/F \Rightarrow (\alpha/F)(\alpha 1/F) = (\alpha/F)(1/F) \\ &\Rightarrow (\alpha\alpha 1)/F = (\alpha 1)/F \Rightarrow \alpha\alpha 1 \in F \Rightarrow \alpha 1 \leqslant \alpha\alpha 1. \end{aligned}$$

The converse inequality depends on Löb's Theorem.

$$\begin{aligned} \alpha 1 \in F &\Rightarrow \alpha 1/F = 1/F \Rightarrow (\gamma\alpha 1)/F = \gamma 1/F \Rightarrow \gamma\alpha 1 \to \gamma 1 \in F \\ &\Rightarrow \alpha 1 \leqslant \gamma\alpha 1 \to \gamma 1 \Rightarrow \gamma\alpha 1 \leqslant \alpha 1 \to \gamma 1 \Rightarrow \tau(\gamma\alpha 1) \leqslant \tau(\alpha 1 \to \gamma 1) \\ &\Rightarrow \alpha\alpha 1 \leqslant \tau\gamma 1 = \alpha 1, \quad \text{by Löb's Theorem.} \end{aligned}$$

Basis. Write $\alpha x = \beta[\tau(\gamma x)]$. Then $\delta = \alpha\beta(1)$. To see this, look at $\tau\gamma\beta(x)$:

$$\begin{aligned} \tau\gamma\beta(1) = \tau\gamma\beta(\tau\gamma\beta(1)) &\Rightarrow \beta[\tau\gamma\beta(1)] = \beta[\tau\gamma\beta(\tau\gamma\beta(1))] \\ &\Rightarrow \alpha\beta(1) = \alpha[\alpha\beta(1)]. \end{aligned}$$

Induction step. Write $\alpha x = \beta[\tau\gamma_1 x, \ldots, \tau\gamma_{k+1} x]$. Letting x_{n+1} be a new variable, we have, by induction hypothesis, a fixed point $\delta_k(x_{n+1})$ to $\beta[\tau\gamma_1 x, \ldots, \tau\gamma_k x, \tau\gamma_{k+1} x_{n+1}]$. The fixed point δ to α is just the fixed point δ to $\delta_k(x_{n+1})$. For,

$$\delta_k(\delta) = \beta[\tau\gamma_1(\delta_k(\delta)), \ldots, \tau\gamma_k(\delta_k(\delta)), \tau\gamma_{k+1}(\delta)].$$

But $\delta = \delta_k(\delta)$, so we can substitute δ for $\delta_k(\delta)$ to get

$$\delta = \delta_k(\delta) = \beta[\tau\gamma_1(\delta), \ldots, \tau\gamma_k(\delta), \tau\gamma_{k+1}(\delta)] = \alpha(\delta). \quad \text{Q.E.D.}$$

Another aspect of the Fixed Point Calculation (1.4) of the preceding section is the uniqueness of the fixed point $a = (\tau a)'$. As Bernardi [5] and de Jongh (cf. [42]) showed, this too generalises.

2.7. EXERCISE (UNIQUENESS OF FIXED POINTS). Let $(A; \tau)$ be a DA, $\alpha(x)$ a polynomial over $(A; \tau)$ in which each occurrence of x lies within the scope of a τ. Then α has a unique fixed point in A.

Both Bernardi and de Jongh actually proved an equationally expressible form of the Uniqueness Theorem: Letting $\tau^+ x = x \cdot \tau x$, they show

$$(1) \qquad \tau^+(x \leftrightarrow \alpha x) \cdot \tau^+(y \leftrightarrow \alpha y) \leqslant x \leftrightarrow y.$$

[Bernardi's proof is a clever application of Löb's Theorem and Lemma 2.5, and the reader might like to try his hand at reconstructing the argument. De Jongh's proof is based on the representation theory for DA's, which we shall shortly describe. To see the equivalence of (1) with uniqueness, (i) assume (1) and note the equivalence, for $z = x, y$ of $z = \alpha z$ with $z \leftrightarrow \alpha z = 1$, and so with $\tau^+(z \leftrightarrow \alpha z) = 1$, to conclude uniqueness; and (ii) appeal to uniqueness in $(A/F; \tau/F)$, where F is the principal τ-filter generated by $\tau^+(x \leftrightarrow \alpha x) \cdot \tau^+(y \leftrightarrow \alpha y)$, to conclude (1).]

The most natural proof of Exercise 2.7 for the present paper is the following: §4, below, contains two proofs of the uniqueness of fixed points for finite FPA's, hence for finite DA's. Since for DA's this uniqueness is equationally expressible, and since the finite DA's are equationally generic for the class of all DA's, uniqueness holds for DA's in general.

3. Representation theory for diagonalisable algebras. As noted by Magari [22], one can read a representation theory for DA's directly off Halmos' representation theory [15] for monadic algebras—this representation theory itself being reducible to Stone's representation theory [53] for boolean algebras. As this representaton theory is vital both in establishing the equational genericity of $(A_{PA}; \tau)$ in the variety of DA's and in proving the main theorem of the present paper, it merits some review here.

3.1. DEFINITION. Let A be a boolean algebra. A subset $F \subseteq A$ is an *ultrafilter* of A if it satisfies

(i) $0 \notin F$,

(ii) $\forall x, y \in A, x, y \in F \Rightarrow x \cdot y \in F$,

(iii) $\forall x, y \in A, x \in F \,\&\, x \leqslant y \Rightarrow y \in F$, and

(iv) $\forall x \in A, x \in F$ or $x' \in F$.

The basic existence result is the following.

3.2. LEMMA. *Let A be a boolean algebra and $F_0 \subseteq A$ a filter (i.e. a subset of A satisfying* 3.1(i)–(iii)). *Then A possesses an ultrafilter $F \supseteq F_0$.*

The proof is a routine application of Zorn's Lemma.

One often applies Lemma 3.2 to sets with the *finite intersection property*—subsets $G \subseteq A$ such that, for any $x_0, \ldots, x_n \in G$, $x_0 \cdot \ldots \cdot x_n \neq 0$. Each such G generates a filter, $F_0 = \{y \in A: \exists x_0 \ldots x_n \in G(x_0 \cdot \ldots \cdot x_n \leqslant y)\}$, which extends by 3.2 to an ultrafilter.

In simplest terms, the Stone Representation Theorem is the following.

3.3. THEOREM (STONE REPRESENTATION THEOREM). *Let A be a boolean algebra and $X = \{F: F \text{ is an ultrafilter of } A\}$. The function*

$$f: x \mapsto \{F: F \in X \& x \in F\}$$

is an isomorphic embedding of A into $P(X)$, i.e. f is one-one and
(i) $f(x \cdot y) = fx \cap fy$,
(ii) $f(x') = X - fx$.

PROOF. Let us first see that f is a homomorphism.

(i) $f(x \cdot y) = \{F: x \cdot y \in F\} \subseteq \{F: x \in F\} \cap \{F: y \in F\}$, by 3.1(iii) since $x \cdot y \leqslant x, y$. The converse inclusion follows from 3.1(ii).

(ii) By 3.1.iv, $fx \cup fx' = X$. By 3.1(i)–(ii), $fx \cap fx' = \varnothing$. It follows that $fx' = X - fx$.

Finally, to see that f is an isomorphic embedding, note that, if $x \neq 0$, then $[x, 1] = \{y \in A: x \leqslant y\}$ is a filter. By Lemma 3.2, there is some $[x, 1] \subseteq F \in X$, i.e. some $F \in fx$. Thus, the kernel of f is $\{0\}$ and f is one-one. Q.E.D.

How does this representation treat τ—i.e. what is the relation between $f(\tau x)$ and fx? Well, it turns out that this relation is clarified by a special type of binary relation on X dual to τ. Why a binary relation? To be honest, I don't know how to motivate this fact algebraically. Logically, I can explain it in terms of modal logic: As far back as Kant mathematical truth, which equals mathematical provability, was hailed as necessary truth. Thus arises the idea of interpreting arithmetic provability, $\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi \urcorner)$, i.e. τx, as necessity, i.e. the modal logician's $\Box p$. If we think of the elements $a \in A$ as sentences φ about the world, the ultrafilters F of A constitute "possible worlds" insofar as they, by exhaustively deciding all assertions, give all complete possible descriptions of the world. Now, depending on one's exact notion of necessity, the matter of what worlds are possible might depend on the given world one is in. (Think, e.g., of future possible worlds.) Thus, one has an *accessibility* relation R determining which worlds are accessible to which: $F R G$ means G is a possible world as far as F is concerned. As to the relation between R and $\Box$, note that $\Box p$ is to assert the truth of p in all possible worlds:

$$F \text{ makes } \Box p \text{ true} \Leftrightarrow \forall G[F R G \Rightarrow G \text{ makes } p \text{ true}].$$

Thus, τ will be dual to a binary relation R.

3.4. DEFINITIONS. Let A be a boolean algebra and X the set of its ultrafilters.
(i) For $\tau: A \to A$, define R_τ by

$$F R_\tau G \Leftrightarrow \forall x \in A(\tau x \in F \Rightarrow x \in G).$$

(ii) For $R \subseteq X \times X$, define τ_R by

$$f(\tau_R x) = \{F: \forall G(F R G \Rightarrow x \in G)\} = X - R^{-1}(fx'),$$

where f is the embedding of Theorem 3.3.

These explain the connections we will establish between the accessibility relation R and the modal operator τ. We already know the key properties of the τ-operators. What are the key properties of the accessibility relations R?

3.5. DEFINITIONS. Let A be a boolean algebra, X the set of its ultrafilters, and $R \subseteq X \times X$ a binary relation on X.

(i) R is *boolean* if the inverse function R^{-1}: $P(X) \to P(X)$, defined by $R^{-1}V = \{F \in X: \exists G \in V(FRG)\}$ for $V \subseteq X$, maps elements of A (i.e., their images fx) to elements of A (again, their images fy).

(ii) R is *relatively reverse well-founded* if, for each $0 \neq x \in A$, fx contains an R-maximal element.

(iii) R is *brrwft* if R is a boolean, relatively reverse well-founded, transitive relation.

Now that we have the crucial brrwft property, we can state the representation theorem.

3.6. THEOREM. *Let A be a boolean algebra and X its set of ultrafilters. Let further τ: $A \to A$ and $R \subseteq X \times X$.*

(i) *If τ is a τ-operator, R_τ is brrwft.*

(ii) *If R is brrwft, τ_R is a τ-operator.*

Moreover, under the assumptions that τ is a τ-operator and R is brrwft,

(iii) $\tau = \tau_{(R_\tau)}$ *and* $R \subseteq R_{(\tau_R)}$.

The mere inclusion of the relation R in its second dual $R_{(\tau_R)}$ is a bit disappointing. In [**15**] Halmos obtains true duality by imposing another condition on R. This condition is topological and lies beyond the scope of the present paper. Happily, we won't need this additional information in the sequel and can safely ignore it.

PROOF OF 3.6. The category theoretic duality works more smoothly if one considers the algebraic dual σ to τ: $\sigma x = (\tau x')'$. While we won't discuss the categorical dual to σ, we will meet σ quite often in the proof.

(i) Let $(A; \tau)$ be diagonalisable. Define R_τ in accordance with 3.4.

R_τ is boolean. Let $x \in A$. Then $R_\tau^{-1}(fx) = f(\sigma x)$, i.e. $R_\tau^{-1}\{G: x \in G\} = \{F: \sigma x \in F\}$.

Half of this is trivial: Let $FR_\tau G$ with $x \in G$. Then $\forall y(\tau y \in F \Rightarrow y \in G)$. If $\sigma x = (\tau x')' \notin F$, then $\tau x' \in F$. But then $x' \in G$ and so $0 = x \cdot x' \in G$, a contradiction. Thus $\sigma x \in F$ and $R_\tau^{-1}(fx) \subseteq f(\sigma x)$.

For the converse inclusion, suppose $\sigma x = (\tau x')' \in F$. Let $G_0 = \{x \cdot y: \tau y \in F\}$. It is easily seen that G_0 has the finite intersection property. If $G \supseteq G_0$ is an ultrafilter, then $FR_\tau G$ and $x \in G$, whence $F \in R_\tau^{-1}(fx)$.

R_τ is relatively reverse well founded. Let $0 \neq x \in A$ and suppose fx has no R_τ-maximal element. Thus $R_\tau^{-1}(fx) \supseteq fx$. But $R_\tau^{-1}(fx) = f(\sigma x) = \{G: \sigma x \in G\}$ and so $x \leqslant \sigma x$, i.e. $\tau x' \leqslant x'$. Löb's Theorem then yields $x' = 1$, i.e. $x = 0$, a contradiction.

R_τ is transitive. Let $FR_\tau GR_\tau H$. Suppose $\tau x \in F$. Since $\tau x \leqslant \tau\tau x$, $\tau\tau x \in F$. But then $\tau x \in G$, whence $x \in H$. Hence $FR_\tau H$.

(ii) Let R be brrwft and τ_R defined in accordance with 3.4. Note first that, since R is boolean, τ_R is indeed defined via 3.4.

$$f(\tau_R x) = \{F: \forall G(FRG \Rightarrow G \in fx)\} = X - R^{-1}(fx').$$

In terms of σ_R, this reads $f(\sigma_R x) = R^{-1}(fx)$.

$\tau_R 1 = 1$, *i.e.* $\sigma_R 0 = 0$. Clearly $R^{-1}\{G: 0 \in G\} = \varnothing$.

$\tau_R x \cdot \tau_R(x \to y) \leqslant \tau_R y$, *i.e.* $\sigma_R x' + \sigma_R(x \cdot y') \geqslant \sigma_R y'$. Note that

$$\sigma_R x' = R^{-1}\{G\colon x \notin G\}, \qquad \sigma_R(x \cdot y') = R^{-1}\{G\colon x \to y \notin G\},$$
$$\sigma_R y' = R^{-1}\{G\colon y \notin G\},$$

where I ignore the f's. Now $\{G\colon y \notin G\} \subseteq \{G\colon x \notin G \text{ or } x \to y \notin G\}$ by modus ponens. But inverse image respects inclusion and union, whence

$$R^{-1}\{G\colon y \notin G\} \subseteq R^{-1}\{G\colon x \notin G\} \cup R^{-1}\{G\colon x \to y \notin G\}.$$

$\tau_R x \leqslant \tau_R \tau_R x$, *i.e.* $\sigma_R x' \geqslant \sigma_R \sigma_R x'$. Dropping the complements, this reads $\sigma_R x \geqslant \sigma_R \sigma_R x$. But

$$f(\sigma_R x) = R^{-1}\{G\colon x \in G\}, \qquad f(\sigma_R \sigma_R x) = R^{-1}R^{-1}\{G\colon x \in G\},$$

and R is transitive: $R^{-1}R^{-1}\{G\colon x \in G\} \subseteq R^{-1}\{G\colon x \in G\}$.

$\tau_R(\tau_R x \to x) \leqslant \tau_R x$. As the reader can easily verify, it suffices to show $\tau_R x \leqslant x \Rightarrow x = 1$. In terms of σ_R this reads $\sigma_R x \geqslant x \Rightarrow x = 0$. Suppose $x \leqslant \sigma_R x$, i.e. $\{F\colon x \in F\} \subseteq R^{-1}\{G\colon x \in G\}$. If $x \neq 0$, $\{F\colon x \in F\}$ is not empty and, by relative reverse well-foundedness, contains a maximal element F_0. By this maximality, $F_0 \notin R^{-1}\{G\colon x \in G\}$, a contradiction. Thus $x = 0$.

(iii) I am getting tired of this proof and so shall only prove one of the two remaining assertions, namely $\tau = \tau_{(R_\tau)}$. Note that it suffices to show: $\forall F \in X[\tau x \in F \Leftrightarrow \tau_{(R_\tau)} x \in F]$.

Let $F \in f(\tau x)$. Let G be given such that $x' \in G$. Then $\sim(F R_\tau G)$ and so $F \in X - R_\tau^{-1}(fx') = f(\tau_{(R_\tau)} x)$.

Let $F \notin f(\tau x)$. As earlier, let $G \in X$ extend $G_0 = \{x' \cdot y\colon \tau y \in F\}$. Clearly $F R_\tau G$, i.e. $F \in R_\tau^{-1}(fx')$. Thus $F \notin f(\tau_{(R_\tau)} x)$. Q.E.D.

Now that we have completed all of this, let us see where it got us. It seems to have led us astray from FPA's in general (Remember them?) and seems also to have given the lie to my promise to discuss *finite* algebras in these opening sections. Well, appearances are often deceiving. I have discussed the representation theory with the finite algebras in mind: For one thing, we can read the finite representation theory off the general one and apply it. For another, we can use the general representation theory to show the class of finite DA's to be equationally generic for the full variety of diagonalisable algebras. Since expounding the general representation theory does not require significantly more effort than expounding the finite theory, this last application justifies our infinitary digression.

Our first task is to draw conclusions about the finite representation theory.

3.7 THEOREM. *Let A be a finite boolean algebra, $X = \{a_0, \ldots, a_n\}$ the set of atoms of A. There is a canonical one-one correspondence between maps $\tau\colon A \to A$ making $(A; \tau)$ diagonalisable and strict partial orderings $\prec$ of X. Moreover, the correspondence of τ with $\prec$ satisfies*

$$a_i \prec a_j \Leftrightarrow \forall x \in A(a_i \leqslant \tau x \Rightarrow a_j \leqslant x),$$
$$\tau x = \sum \{a_i \in X\colon \forall a_j \in X(a_i \prec a_j \Rightarrow a_j \leqslant x)\}$$

where the ordering $\leqslant$ is that of the boolean algebra.

PROOF. But for the fact that the correspondence is one-one (verification of which I leave to the reader), this follows directly from Theorem 3.6. For, (i) the ultrafilters F of a finite boolean algebra A are exactly the principal ultrafilters $[a_i, 1]$ generated by the atoms $a_i \in A$; and, consequently, (ii) the function f of that Theorem is onto: If $Y = \{F_{i_0}, \ldots, F_{i_k}\}$ is a set of ultrafilters F_{i_j} generated by atoms a_{i_j}, then $Y = f(a_{i_0} + \cdots + a_{i_k})$. The former fact allows us to identify the present set X of atoms with the earlier set X of ultrafilters. The second fact tells us that all binary relations R on X are boolean. It only remains to note that the relatively reverse well-founded transitive relations on a finite set are just the strict partial orderings of that set. Q.E.D.

This representation theory, known as Kripke semantics to modal logicians, is of supreme importance in understanding diagonalisable algebras. The transition from the algebra to the partial ordering logarithmically reduces the cardinality of the structure (that is, in the *finite* case). Moreover, basic operations on structures tend to be simpler and more easily visualisable in terms of the partial orderings. (Consider, e.g., the fact that the dual in Stone's theory to the cartesian product is the disjoint union of the spaces of ultrafilters.) Thus, Kripke's semantics tend to be more easily exploited and the best results are usually proven via the representation theory rather than directly via the algebras.

As a first application of the representation theory, I shall present a proof that the finite DA's are equationally generic in the variety of DA's, i.e. that the finite DA's make no new equations valid. This is an oft-proven result. It was proven by modal logicians before the connection with self-reference was emphasised (e.g. [**39**]); it was reproven by a universal algebraist unaware of the earlier work on the modal logic [**4**]; and it was reproven by other logicians when they too became interested in the modal logic (e.g. [**50**]). The method of proof is routine modal-logically speaking; algebraically it might appear clever. There are two versions of the proof. In the version to be given here, one starts with an infinite DA in which an equation is invalid and "extracts" a finite DA from it in which the equation is also invalid. In the other version, one builds the finite representation directly. [The second proof does not depend on one's having made a successful analysis of the infinite representation theory (in logical terms: It doesn't presuppose completeness of the modal logic) and this fact might be useful. I refer the reader to [**50**] for this proof and to [**42**] for a modification and additional application.]

ONE LAST POINT. I shall prove a somewhat stronger result than mere equational genericity.

3.8. THEOREM. *The universal theory of the class of finite* DA*'s coincides with that of all* DA*'s. Moreover, this common universal fragment of the two theories is decidable.*

I will comment on disagreement and undecidability after having proven the theorem. The task of proving this theorem requires us to think logically—or, at least, linguistically. That is, it is most convenient to formally view polynomials as syntactic objects to be semantically interpreted.

Formally, polynomials are terms constructed according to the following rules:

(i) variables $x_1, x_2, \ldots$ are terms;

(ii) if s, t are terms, so are $(s \cdot t)$, (s') and (τs). A term s is a subterm of t if s appears during the above generation of t. In other words, the relation "s is a subterm of t" is inductively defined by

(i) s is a subterm of s,

(ii) if s is a subterm of t or of u, then s is a subterm of $(t \cdot u)$,

(iii) if s is a subterm of t, then s is a subterm of (t') and (τt). The excessive use of parentheses in the first of these inductive definitions, an excess in which we shall not indulge, guarantees the unique readability of terms. Thus, functions on terms can be inductively defined. In particular, their valuations in DA's are well-defined: Let $(A; \tau)$ be a DA and let g be any assignment of values in A to the variables. The g-valuation in A of a term t, written $[t]_A$ (g is suppressed), is inductively defined by

$$[x_i]_A = gx_i, \qquad [s \cdot t]_A = [s]_A \cdot [t]_A, \qquad [s']_A = ([s]_A)',$$
$$[\tau s]_A = \tau([s]_A).$$

With all of this, we are in position to prove the theorem.

PROOF OF THEOREM 3.8. Note that any universal assertion true in all DA's is true in all finite DA's. The converse is proven contrapositively.

Note first that any equation $s = t$ can be rewritten $s \leftrightarrow t = 1$, any inequation $s \neq t$ can be rewritten $s \leftrightarrow t \neq 1$, and inequalities can similarly be rewritten (e.g. $s \leqslant t$ as $s \to t = 1$). Combining this observation with our ability to put boolean formulae into conjunctive normal form, we see that a universal assertion can be put into the form

$$\forall x_0 \ldots x_n \left[\bigwedge_i \left(\bigvee_j t_{ij} = 1 \vee \bigvee_k t_{ik} \neq 1 \right) \right], \tag{1}$$

(where the i's, j's and k's range over disjoint finite sets). Since universal quantifiers distribute over conjunction, (1) can be rewritten as

$$\bigwedge_i \forall x_0 \ldots x_n \left(\bigvee_j t_{ij} = 1 \vee \bigvee_k t_{ik} \neq 1 \right). \tag{2}$$

As such a conjunction is valid in either class of algebras (i.e. finite vs. general) iff each conjunct is, we finally see that it suffices to show the preservation from the finite to the general case of the validity of assertions of the form

$$\forall x_0 \ldots x_n \left(\bigvee_i t_i = 1 \vee \bigvee_j t_j \neq 1 \right). \tag{3}$$

Now we can begin the proof in earnest. Suppose $(A; \tau)$ is a DA in which (3) fails. Then there is an interpretation $[x_0]_A, \ldots, [x_n]_A$ of the variables $x_0, \ldots, x_n$ such that

(i) for each t_i there is an ultrafilter F_i on A with $[t_i]_A \notin F_i$ and

(ii) for each t_j and each ultrafilter F on A $[t_j]_A \in F$.

While both of these facts will be used later, at the moment we will only use the first. We use the filters F_i to generate a finite rootless tree $(Y, \prec)$, which will in turn generate a finite DA.

We begin the construction of $(Y, \prec)$ by putting the unit sequences $\langle F_i \rangle$ into Y. This is stage 1.

Let X be the set of ultrafilters on A, R the binary relation on X dual to τ. Define

$$S = \{s\colon s \text{ is a subterm of some } t_i \text{ or } t_j\}.$$

Stage $k+1$ of the construction is now simple: We let $p = \langle F_0, \ldots, F_{k-1} \rangle$ arise at stage k. The one-step extensions $q = \langle F_0, \ldots, F_{k-1}, F_k \rangle$ of p added to Y at stage $k+1$ are determined as follows. For each $\tau s \in S$ such that $[\tau s]_A \notin F_{k-1}$, choose an ultrafilter $F \in X$ such that

(i) $F_{k-1} R F$ and

(ii) $[\tau s]_A \in F \& [s]_A \notin F$.

[This can be done by Löb's Theorem. Note that F is an R-maximal element of $\{G\colon [s]_A \notin G \& F_{k-1} R G\}$.] For the chosen F, we put $q = \langle F_0, \ldots, F_{k-1}, F \rangle$ into Y.

We continue this procedure until it can no longer be followed. Note first that at any stage only finitely many new sequences are added to Y—basically because we add at most one sequence q for each pair $\langle p, \tau s \rangle$, where p is already given and $\tau s \in S$. Second, note that the process stops after at most card(S) stages. For, when $\tau s \in S$ is considered for p at stage k, one extends to $q = \langle F_0, \ldots, F_k \rangle$ for which $[\tau s]_A \in F_k$. Thus, for any $G \in X$ satisfying $F_k R G$, $[\tau s]_A \in G$ and τs will not be considered relative to any extension of p. After card(S) steps we will have run out of terms τs to consider. Thus, if we order Y by proper extension, i.e. $p \prec q$ iff p is a proper initial segment of q, then Claim 1 holds:

CLAIM 1. $(Y, \prec)$ is a finite strictly partially ordered set.

By Theorem 3.7, this means that there is a τ-operator defined on the power set algebra $P(Y)$. We define a valuation on this algebra—call it C—as follows:

$$[x]_C = \begin{cases} \{p \in Y\colon [x]_A \in F_k\}, & \text{where } p = \langle F_0, \ldots, F_k \rangle, k \geqslant 0, \\ & \text{and } x \in \{x_0, \ldots, x_n\}, \\ \varnothing, & \text{if } x \notin \{x_0, \ldots, x_n\}. \end{cases}$$

CLAIM 2. For all $s \in S$ and all $p = \langle F_0, \ldots, F_k \rangle$ $(k \geqslant 0)$,

$$p \in [s]_C \Leftrightarrow [s]_A \in F_k.$$

The proof of this claim is an induction on the generation of s. For variables x_i, the claim holds by definition.

If s is $t \cdot u$, then

$$\begin{aligned} p \in [t \cdot u]_C &\Leftrightarrow p \in [t]_C \& p \in [u]_C \\ &\Leftrightarrow [t]_A \in F_k \& [u]_A \in F_k \Leftrightarrow [t \cdot u]_A \in F_k. \end{aligned}$$

If s is t', Lemma 3.2 applies.

$$p \in [t']_C \Leftrightarrow p \notin [t]_C \Leftrightarrow [t]_A \notin F_k \Leftrightarrow [t']_A \in F_k.$$

Finally, suppose s is τt;

$$p \in [\tau t]_C \Leftrightarrow \forall q \succ p(q \in [t]_C) \Leftrightarrow \forall q \succ p([t]_A \in F_{k+1}),$$

writing $q = \langle F_0, \ldots, F_k, F_{k+1} \rangle$. Continuing,

$$\Rightarrow [\tau t]_A \in F_k,$$

or else we would have chosen a $q = \langle F_0, \ldots, F_k, F_{k+1} \rangle$ at stage $k + 2$ such that $F_k \, R \, F_{k+1}$ and $[t]_A \notin F_{k+1}$.

Conversely,

$$p \notin [\tau t]_C \Leftrightarrow \exists q \succ p(q \notin [t]_C)$$
$$\Rightarrow \exists F_{k+1}(F_k \, R \, F_{k+1} \& [t]_A \notin F_{k+1}) \Rightarrow [\tau t]_A \notin F_k.$$

This completes the proof of the claim. As immediate consequences, we note

(i) $\langle F_i \rangle \notin [t_i]_C$ for the initially mentioned F_i and

(ii) $p \in [t_j]_C$ for all $p \in Y$ and all t_j. Thus, the universal assertion (3) fails in the finite DA $(C; \tau)$.

All that remains is the observation that the cardinalities of Y and C are effectively bounded by knowledge of the terms t_i and t_j. Hence, to test the validity of (3) in all DA's it suffices to test this validity in an effectively determined finite set of finite structures. The universal fragment of the theory is, thus, decidable. Q.E.D.

After any proof it is always a good idea to relax among simple considerations, such as e.g. remarks on how good the result is.

3.9. REMARKS. Theorem 3.8 is best possible in one sense and pretty good in another.

(i) The agreement of the theories of finite and of all DA's doesn't go much further. The finite algebras satisfy new existential assertions. It is not hard, for example, to see that $\exists x(x \neq 1 \& \tau x = 1)$ is valid in all finite DA's, but not in all DA's (in particular, not in the Lindenbaum DA $(A_{\mathrm{PA}}; \tau)$). [SUBREMARK. If we restrict our attention to the purely boolean language, the agreement goes a bit further. Since finitely generated boolean algebras are finite, the $\forall\exists$ fragments of the boolean theories coincide. The $\exists\forall$ fragments, however, do not as the ordering of a finite boolean algebra is never dense while that of an infinite one (e.g. any atomless boolean algebra) can be—and nondensity is $\exists\forall$-expressible:

$$\exists xy \forall z(x < y \& \neg(x < z \& z < y)).]$$

(ii) The decidability result does not extend to the full language for either the theory of all DA's (as shown by Montagna in [**31**]) or the theory of finite DA's (as shown by the author in [**49**]—for the specialist I note that I prove finite inseparability). [SUBREMARK. For the boolean language this is false. Both the theories of boolean algebras and of finite boolean algebras have long been known decidable.]

Now that we are rested, it is time to discuss another result on equational genericity—namely, that of the structure $(A_{PA}; \tau)$.

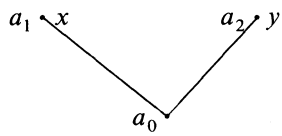
3.10. THEOREM (SOLOVAY'S COMPLETENESS THEOREM). *The* DA $(A_{PA}; \tau)$ *is equationally generic for the variety of* DA*'s.*

Before commenting on the proof, I note that this result is best possible: $(A_{PA}; \tau)$ satisfies variable-free inequations not valid in all DA's. A simple example is $\tau 0 \neq 1$. Additional examples are $\tau\tau 0 \neq 1$, $\tau\tau\tau 0 \neq 1$, etc. Even discounting these, i.e. even restricting our attention to those DA's in which these inequations are true, we find additional universal assertions true in $(A_{PA}; \tau)$—e.g. $\forall x(\tau x = 1 \supset x = 1)$.

Solovay's proof (in [**50**]) of Theorem 3.10 requires a bit too much logic to be presented here; so I shall merely describe an example of the use of his method and thereby give something of the flavour of the proof. Let us take an equation which is not valid in all DA's, say

$$\tau(x + y) \to (\tau x + \tau y) = 1. \tag{4}$$

We search for a finite DA in which this is not valid and look at its representation. For example, there is the simple tree



What this indicates is that we have the 8-element boolean algebra A with atoms a_0, a_1, a_2 ordered as shown, and x, y coincide with the atoms a_1, a_2, respectively. Now $x + y = \{a_1, a_2\}$ and $\tau(x + y) = \{a_1, a_2, a_3\} = 1$. But $\tau x = \{a_1, a_2\}$ and $\tau y = \{a_1, a_2\}$, whence $\tau x + \tau y = \{a_1, a_2\} \not\ni \tau(x + y)$. Solovay next proceeds to simulate this representation within PA by constructing sentences (by self-reference—but, alas, nonextensional self-reference) ψ_0, ψ_1, ψ_2 such that

(i) $\mathrm{PA} \vdash \psi_0 \vee \psi_1 \vee \psi_2$,

(ii) $\mathrm{PA} \vdash \psi_i \supset \bigwedge_{a_i \not< a_j} \mathrm{Pr}_{PA}(\ulcorner \neg\psi_j \urcorner)$, for each $i = 0, 1, 2$,

(iii) $\mathrm{PA} \vdash \psi_i \supset \bigwedge_{a_i < a_j} \neg \mathrm{Pr}_{PA}(\ulcorner \neg\psi_j \urcorner)$, for each $i = 0, 1, 2$.

Solovay shows that, with these relations, the sentences ψ_i simulate the atoms exactly. In particular, from the fact that $a_0 \not\leqslant \tau(a_1 + a_2) \to (\tau a_1 + \tau a_2)$, it follows that

$$\mathrm{PA} \not\vdash \psi_0 \supset \mathrm{Pr}_{PA}(\ulcorner \psi_1 \vee \psi_2 \urcorner) \supset \mathrm{Pr}_{PA}(\ulcorner \psi_1 \urcorner) \vee \mathrm{Pr}_{PA}(\ulcorner \psi_2 \urcorner). \tag{5}$$

In particular,

$$\mathrm{PA} \not\vdash \mathrm{Pr}_{PA}(\ulcorner \psi_1 \vee \psi_2 \urcorner) \supset \mathrm{Pr}_{PA}(\ulcorner \psi_1 \urcorner) \vee \mathrm{Pr}_{PA}(\ulcorner \psi_2 \urcorner),$$

and the equation (4) is not valid in $(A_{PA}; \tau)$. [*Exercise.* Use properties (i)–(iii) to derive (5) directly.]

For the complete proof of Theorem 3.10 I refer the reader to [**50**]. Modifications of the proof and refinements of the result can be found in [**1**, **29**, and **57**]

(and, in the study of nonextensional self-reference, Guaspari and Solovay have also applied the above technique in [**11**]).

As I remarked earler, Solovay's Theorem is best possible with respect to agreement between the theory of all DA's and the theory of the single structure $(A_{\mathrm{PA}}; \tau)$. It says nothing, however, about other possible connections. Far more useful than Theorem 3.10 is another result Solovay proved along with this Theorem.

3.11. THEOREM (SOLOVAY'S SECOND COMPLETENESS THEOREM). *Let t be a term, $S = \{s\colon \tau s$ is a subterm of $t\}$. The following are equivalent.*

(i) *Every interpretation of t within the language of arithmetic is true.*

(ii) $\prod_{s\in S}(\tau s \to s) \leqslant t$ *is valid in* $(A_{\mathrm{PA}}; \tau)$ *and hence in all* DA's.

To illustrate what (i) means, I offer the term

$$t_0(x_0, x_1)\colon \quad \tau x_0 \cdot (\tau x_1 \to x_1).$$

Assertion (i) about t_0 then claims that, for every choice of arithmetic sentences φ_0, φ_1, the sentence $t_0(\varphi_0, \varphi_1)$ (i.e. $\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_0 \urcorner) \wedge (\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_1 \urcorner) \supset \varphi_1)$) is *true* in the nonnegative integers (as opposed to being provable in PA—which reads "$t = 1$ is valid in $(A_{\mathrm{PA}}; \tau)$"). For the term t_0 cited, this assertion is false; for the term t_1,

$$t_1(x_0, x_1)\colon \quad \tau x_0 + (\tau x_1 \to x_1),$$

however, the corresponding assertion is true. Thus, by (ii), for some φ_0, φ_1,

$$\mathrm{PA} \not\vdash \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_0 \urcorner) \supset \varphi_0\big] \wedge \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_1 \urcorner) \supset \varphi_1 \urcorner\big] \supset \mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_0 \urcorner) \wedge \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_1 \urcorner) \supset \varphi_1\big],$$

but, for all φ_0, φ_1,

$$\mathrm{PA} \vdash \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_0 \urcorner) \supset \varphi_0\big] \wedge \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_1 \urcorner) \supset \varphi_1\big] \supset \mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_0 \urcorner) \vee \big[\mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi_1 \urcorner) \supset \varphi_1\big].$$

I will not discuss the proof of Theorem 3.11 here. I note simply that (i) the proof is a modification (better: an extension) of that of Theorem 3.10; (ii) this theorem is far and away more useful than Theorem 3.10, for example, in that it allows one to draw many refined conclusions about the nature of undecidability in PA by simple considerations of propositional modal logic; and (iii) this theorem has as an immediate corollary the following:

3.12. COROLLARY. *The universal theory of the structure $(A_{\mathrm{PA}}; \tau)$ is decidable.*

PROOF. Consider the old universal assertion (3), rewritten here in the form

$$\forall x_0 \ldots x_n \Big(\bigwedge_j t_j = 1 \supset \bigvee_i t_i = 1 \Big). \tag{6}$$

Suppose first that (6) is true in $(A_{\mathrm{PA}}; \tau)$. Then, for any choice of sentences $\varphi_0, \ldots, \varphi_n$,

$$\vDash \bigwedge_j \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_j(\varphi_0, \ldots, \varphi_n)\urcorner\right) \Rightarrow \vDash \bigvee_i \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_i(\varphi_0, \ldots, \varphi_n)\urcorner\right),$$

where "$\vDash$" denotes truth in the natural numbers. Since we are dealing with truth rather than provability, this implication yields

$$\vDash \bigwedge_j \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_j(\varphi_0, \ldots, \varphi_n)\urcorner\right) \supset \bigvee_i \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_i(\varphi_0, \ldots, \varphi_n)\urcorner\right).$$

Since $\varphi_0, \ldots, \varphi_n$ are arbitrary, Theorem 3.11 and a little algebra apply to yield the validity in all DA's of

$$\forall x_0 \ldots x_n \left[\prod_k (\tau s_k \to s_k) \cdot \prod_j (\tau t_j) \leqslant \sum_i (\tau t_i)\right], \tag{7}$$

for suitable terms s_k.

Conversely, the validity of (7) in all DA's entails that of (6) in $(A_{\mathrm{PA}}; \tau)$: Pick $\varphi_0, \ldots, \varphi_n$ arbitrarily. Suppose, for each j, $t_j(\varphi_0, \ldots, \varphi_n) = 1$ in $(A_{\mathrm{PA}}; \tau)$, i.e.

$$\vDash \bigwedge_j \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_j(\varphi_0, \ldots, \varphi_n)\urcorner\right). \tag{8}$$

But it is evident that

$$\vDash \bigwedge_k \left[\mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner s_k(\varphi_0, \ldots, \varphi_n)\urcorner\right) \supset s_k(\varphi_0, \ldots, \varphi_n)\right]. \tag{9}$$

Now the validity of (7) in $(A_{\mathrm{PA}}; \tau)$ means the provability of its interpretations in PA, whence the truth of this interpretation in the natural numbers. With (8) and (9), this yields

$$\vDash \bigvee_i \mathrm{Pr}_{\mathrm{PA}}\left(\ulcorner t_i(\varphi_0, \ldots, \varphi_n)\urcorner\right).$$

Pick a true disjunct $\mathrm{Pr}_{\mathrm{PA}}(\ulcorner t_{i_0}(\varphi_0, \ldots, \varphi_n)\urcorner)$. This true statement asserts simply that

$$t_{i_0}(\varphi_0, \ldots, \varphi_n) = 1$$

in $(A_{\mathrm{PA}}; \tau)$. As $\varphi_0, \ldots, \varphi_n$ were arbitrary, we conclude (6).

The equivalence between the validity of (6) in $(A_{\mathrm{PA}}; \tau)$ and the validity of its effectively obtained equational equivalent (7) in $(A_{\mathrm{PA}}; \tau)$ and hence in all DA's reduces the decision problem for universal sentences over $(A_{\mathrm{PA}}; \tau)$ to the decidable corresponding problem over all DA's. Q.E.D.

I note that it is an open problem whether or not the full first-order theory of the structure $(A_{\mathrm{PA}}; \tau)$ is decidable.

I will let the reader draw his own conclusions about the sorts of applications to the study of formal arithmetic that these completeness and decidability results offer.

I close this section with a couple of direct applications of the finite representation theory to the study of DA's.

3.13. EXERCISE. Give de Jongh's proof of the uniqueness of fixed points: Given a finite partially ordered set $(P, \prec)$, show by induction from the top

down that, for any appropriate polynomial $\alpha(x)$, there is a unique way of deciding $a \leqslant x$ (in the boolean sense) for atoms $a \in P$ so that $a \leqslant (x \leftrightarrow \alpha x) \cdot \tau(x \leftrightarrow \alpha x)$.

3.14. EXERCISE. Show the class of equations valid in all DA's to be closed under the following rules of inference:

(i) $\tau t = 1 \Rightarrow t = 1$;

(ii) $\tau s \leqslant \tau t \Rightarrow s \cdot \tau s \leqslant t$;

(iii) $\tau(s + t) = 1 \Rightarrow \tau s = 1$ or $\tau t = 1$, for any terms s, t. [HINT. Consider, for example, (i). Let $(P \prec)$ be a representation for a finite algebra in which $t = 1$ is not valid. Let $(P', \prec')$ arise from $(P, \prec)$ by the adjunction of a new minimum element. Show $\tau t = 1$ fails to be valid in the corresponding algebra.]

4. Finite fixed point algebras. As the reader who hasn't completely forgot about FPA's has doubtlessly gathered, the DA's form a very important class of FPA's. As we shall see, the finite DA's constitute the very core of the class of finite FPA's. The present section is devoted to studying a few basic properties of finite FPA's. Among other things, it will include

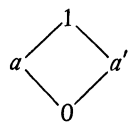
(i) examples of finite FPA's (B, A), including complete catalogues of the B's for $A = 2$ and 4, and a few bounds on card(B) for $A = 8$ (where 2, 4, and 8 are, respectively, the 2-element, 4-element, and 8-element boolean algebras);

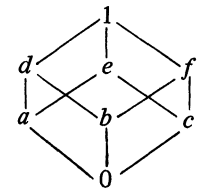
(ii) proofs of the uniqueness of fixed points in finite FPA's; and,

(iii) a minor structural lemma that will be useful later in proving the main theorem of the present paper.

For the sake of the examples, we label the elements of 2, 4, and 8.

2: 

4: 

8: 

4.1. EXAMPLE. The only algebra B making $(B, 2)$ an FPA is the trivial one $B = \{\lambda x.0, \lambda x.1\}$.

[*A word about* λ*-notation.* The distinction between the function f given by an expression fx and its value at an argument x denoted by fx often has to be made. One way of doing this is to let $\lambda x.fx$ denote the function. A constant function $\lambda x.a$ will occasionally be denoted a—i.e. we identify A with the subalgebra of B of constant functions.]

4.2. EXAMPLES. The only algebras B for which $(B, 4)$ is an FPA are the following:

(i) $B_0 = 4$;

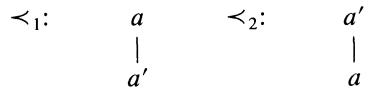
(ii) $B_1 = 4 \cup \{\tau_1, \lambda x.(\tau_1 x)', \lambda x.(a' \cdot \tau_1 x), \lambda x.(a + (\tau_1 x)')\}$;

(iii) $B_2 = 4 \cup \{\tau_2, \lambda x.(\tau_2 x)', \lambda x.(a \cdot \tau_2 x), \lambda x.(a' + (\tau_2 x)')\}$,

where τ_1, τ_2 are the τ-operators

x	0	a	a'	1
$\tau_1 x$	a	1	a	1
$\tau_2 x$	a'	a'	1	1

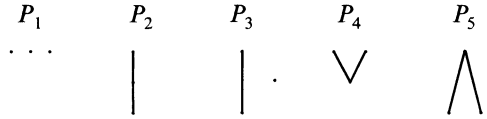
with respective dual orderings



[Note that B_0 has the τ-operator $\tau_0 = \lambda x.1$ determined by the empty ordering.]

The assertion of Example 4.1 is trivial. The positive assertions in Example 4.2 that B_0, B_1, and B_2 yield FPA's on $A = 4$ are simple matters of (i) checking each of the given functions to have a fixed point, and (ii) checking each list B_i of functions to be closed under the boolean operations and composition. The fact that the functions τ_0, τ_1, and τ_2 exhaust all τ-operators on 4 (and thus that B_0, B_1, and B_2 yield all FPA's $(B, 4)$ that are DA's) follows quickly from the representation theory. The fact that the algebras B_0, B_1, and B_2 are the only algebras B for which $(B, 4)$ is an FPA can be verified computationally in a few pages. Shortly we will be in position to see this with a minimal amount of computation.

4.3. EXAMPLES. There are 19 τ-operators on 8 corresponding to the assignments of the atoms a, b, c of 8 to nodes of the following partial order types



[There are 1, 6, 6, 3, and 3 orderings of types P_1, P_2, P_3, P_4, and P_5, respectively.]

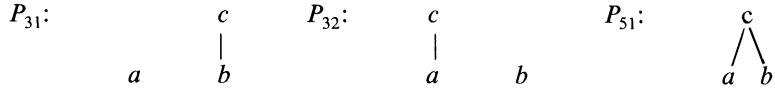
A complete catalogue of FPA's on 8 is within the grasp of anyone with access to a computer, but is too complicated (as far as I am concerned) for mere hand calculation. Three things complicate the picture.

(i) Viewed as FPA's, distinct DA's can coincide.

(ii) Proper subalgebras of DA's, which are not themselves DA's, can be FPA's.

(iii) The FPA's can be rather large for paper and pencil.

4.3. EXAMPLES (CONTINUED). (i) The FPA's given by polynomials in the τ-operators of the following partial orderings coincide



To verify this, note first that the corresponding τ-operators are defined by

x	$0, a, b, d$	$c, e, f, 1$
$\tau_{31}x$	e	1
$\tau_{32}x$	f	1
$\tau_{51}x$	c	1

These operators are interdefinable, e.g.

$$\tau_{32} = \lambda x.(f + \tau_{51}x), \qquad \tau_{51} = \lambda x.(c \cdot \tau_{31}x).$$

If $\alpha\colon A \to A$ and $\alpha \in B$ for some FPA (B, A), we let $A[\alpha]$ denote the smallest algebra B such that (B, A) is an FPA and $\alpha \in B$; i.e. $A[\alpha]$ is the algebra generated from α and the constants in A by composition and the boolean operations. [*Note*. We only need the initial assumption that $\alpha \in B$ for some FPA (B, A) to conclude that $(A[\alpha], A)$ is an FPA, not to define $A[\alpha]$.]

4.3. EXAMPLES (CONTINUED FURTHER). Let $A = 8$.

(i) Let τ_{21} be the τ-operator

x	$0, a, b, d$	c, e	$f, 1$
$\tau_{21}x$	c	f	1

determined by the ordering

P_{21}:

The algebra $A[\tau_{21}]$ is a proper subalgebra of the DA generated by τ_{21}, as is the algebra $A[\lambda x.\tau_{21}x']$.

(ii) The operator α_0 defined by

x	$0, a$	$b, c, d, e, f, 1$
$\alpha_0 x$	1	f

is of the form $\lambda x.\tau_{41}x'$ for some τ-operator τ_{41} on P_4. Thus, $(A[\alpha_0], A)$ is a subdiagonalisable FPA. However, this algebra contains no nonconstant τ-operator.

Part (i) is easily verified: Note that every $\alpha \in A[\tau_{21}]$ maps $0, a, b, d$ to the same element; but $\lambda x.\tau_{21}x'$ maps a to 1 and b to f. Thus $\lambda x.\tau_{21}x' \notin A[\tau_{21}]$. Similarly, $\tau_{21} \notin A[\lambda x.\tau_{21}x']$. Each of these algebras contains nonconstant τ-operators (exercise). That the algebra of part (ii) contains no nonconstant τ-operator is seen by inspection—the algebra has only 16 elements, which the reader can easily generate.

If one picks an α and simply tries to generate $A[\alpha]$ an element at a time, he will usually get frustrated. A systematic approach is time-consuming and boring; while an "inspired" approach can quickly lead to conjectures on the forms of the functions $\beta \in A[\alpha]$, but also to minor difficulty in obtaining a specific desired β. The reader is urged to judge for himself by generating the 64 elements of $A[\lambda x.\tau_{21}x']$, the 64 elements of $A[\tau_{21}]$, and the 128 elements of the full DA (B, A) generated by τ_{21}. [Of course, the job is not finished until all functions have been tested for fixed points and all closure conditions have been verified!]

I said earlier that the cardinality of the algebra B can be quite large and I have just noted an example of a B making $(B, 8)$ an FPA with the cardinality of B being 128. For $A = 8$ this is the largest example I have—roughly my limit for paper and pencil. Moreover, for reasons to appear in the next section, this example and its 5 isomorphs (corresponding to the other 5 orderings on P_2) are

the largest such B's that I expect to exist. [Similarly, there are 24 B's of cardinality 32, 768 (which I suspect to be of maximum cardinality) on $A = 16$.]

Let me give some upper bounds on the cardinality of an algebra B making $(B, 8)$ an FPA. These give useful measures of our knowledge—the lower the bound, the better our grasp of the situation.

For convenience, from here on I will denote the cardinality of B, formerly card(B), by $|B|$.

The grossest bound is given by noting only that $B \subseteq A^A$.

4.4. FIRST BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 16{,}777{,}216$.

For $|B| \leqslant 8^8 = 16{,}777{,}216$.

If we bother to note that the definition of a fixed point algebra requires each $\alpha \in B$ to have a fixed point, we can ignore those $\alpha \in A^A$ having no fixed points. There are $7^8 = 5{,}764{,}801$ of these.

4.5. SECOND BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 11{,}012{,}415$.

If we recall that, B being a boolean algebra, $|B|$ must be a power of 2, we can reduce this a bit by taking the greatest power of 2 not exceeding the cited bound.

4.6. IMPROVED SECOND BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 8{,}388{,}608$.

Compared with the expected $|B| \leqslant 128$, this shows we don't know much yet other than the definition. Let's see—I have already announced that I will show every $\alpha \in B$ to have a unique fixed point. Does this knowledge help? Note that $|\{\alpha\colon 8 \to 8\colon \alpha \text{ has a unique fixed point}\}| = 8 \cdot 7^7 = 6{,}588{,}344$. Looking at the greatest power of 2 not exceeding this yields a third bound.

4.7. THIRD BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 4{,}194{,}304$.

This is still abysmal. It is time to introduce some machinery.

4.8. LEMMA. *Let A be a boolean algebra, $a, b \in A$ with $a < b$. The interval*

$$[a, b] = \{x \in A\colon a \leqslant x \leqslant b\}$$

is a boolean algebra in its own right, where

(i) *the* 0 *and* 1 *of* $[a, b]$ *are* a *and* b, *respectively,*

(ii) *the lattice operations* (*i.e.* $+, \cdot$) *of* $[a, b]$ *are those inherited from* A, *and*

(iii) *the complement is relativised,* $\nu x = b \cdot (a + x')$.

The only portion of this lemma needing verification is that ν defines a complement in $[a, b]$, i.e. that

$$x + \nu x = b, \qquad x \cdot \nu x = a,$$

for $x \in [a, b]$. These are simple calculations which the reader can either carry out or look up in any text on boolean algebras.

4.9. DEFINITION. Let (B, A) be an FPA, and $a, b \in A$ with $a < b$. For each $\alpha \in B$, define its "retraction" α^r to $[a, b]$ by

$$\alpha^r x = b \cdot (a + \alpha x).$$

We ambiguously denote both the function from B to $[a, b]$ just defined and its restriction mapping $[a, b]$ to $[a, b]$ by α^r. In terms of these restrictions, we define

$$B^r = \{\alpha^r\colon \alpha \in B\}.$$

Further, we define operations $+, \cdot, '$ pointwise (using ν for $'$) and call the resulting structure $(B^r, [a, b])$ the algebra induced on $[a, b]$ by (B, A), or more simply the *induced algebra*.

4.10. THEOREM. *Let* (B, A) *be an* FPA, $a, b \in A$ *with* $a < b$. *The induced algebra* $(B^r, [a, b])$ *is an* FPA.

PROOF. The fact that the map $x \mapsto b \cdot (a + x)$ is a homomorphism of A onto the algebra $[a, b]$ readily yields the closure of B^r under the boolean operations —hence the fact that B^r is a boolean algebra.

If $c \in [a, b]$ then $b \cdot (a + c) = c$ and $(\lambda x.c)^r$ is the constant function c on $[a, b]$.

Further, B^r is closed under composition

$$\beta^r(\alpha^r x) = b \cdot (a + \beta(\alpha^r x)) = (\lambda x.\beta(\alpha^r x))^r \in B^r,$$

since $\alpha^r \in B$ and B is closed under composition.

Finally, we must see that each $\alpha^r \in B^r$ possesses a fixed point in $[a, b]$. Viewed as an element of B, α^r has a fixed point, say, $c \in A$. But $c = \alpha^r c \in \text{range}(\alpha^r) \subseteq [a, b]$. Q.E.D.

This simple construction is a matter of some consequence—it connects the algebra B operating on A with an algebra B^r operating on a smaller algebra $[a, b]$. This construction allows for an inductive understanding of finite FPA's. For example, analysing which functions from A to A can induce those of B^r allows us to bound $|B|$ in terms of $|B^r|$.

4.11. THEOREM. *Let* (B, A) *be a finite* FPA, $a \in A$ *an atom, and* $(B^r, [a, 1])$ *an induced algebra. Then* $|B| \leqslant \frac{1}{2}|B^r| \cdot (2 \cdot |A|)^{|A|/2}$.

PROOF. Pick $\alpha^r \in B^r$, $c \in A$. There are *two* elements $d_0, d_1 \in A$ for which $a + d_i = \alpha^r c = a + \alpha c$. Hence, there are $2^{|A|/2}$ possible choices of β: $[a, 1] \to [a, 1]$ for which $\beta^r = \alpha^r$. This gives the factor $|B^r| \cdot 2^{|A|/2}$. The factor $|A|^{|A|/2}$ comes from the arbitrary complementary functions β: $[0, a'] \to [0, a']$.

However, we can reduce this bound to the next lower power of 2 by exhibiting a function counted among our possibilities that is definitely not in B. Given α^r, define β by

$$\beta x = \text{some } y \neq x \text{ such that } a + y = \alpha^r x.$$

Note that $\beta^r = \alpha^r$, but that β has no fixed point and hence $\beta \notin B$. Q.E.D.

As a corollary, we get an improved bound for the case $A = 8$.

4.12. FOURTH BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 262{,}144$.

PROOF. First note that, for any atom g, $(B^r, [g, 1])$ is essentially of the form $(B_*, 4)$. But I have already remarked (albeit not proven) that B_* must be one of the three algebras of Examples 4.2. The largest of these have cardinality 8. Thus, $|B| \leqslant \frac{1}{2} \cdot 8 \cdot (2 \cdot 8)^{8/2} = 262{,}144$. Q.E.D.

Two things are to be noted about this proof. First, it gives a better bound than that achieved by appeal to the uniqueness of fixed points—which uniqueness we have not yet proved. Second, using the same technique to bound $|B|$ when $(B, 4)$ is an FPA would yield $|B| \leqslant 128$—not $|B| \leqslant 8$ as we used.

So, by analogy, the bound given for $A = 8$ is likely still to be too large. Indeed, later in the section we will lower this bound to 1024.

Using the induced algebra, we can do more than estimate the cardinalities of algebras. For one thing, we can prove the uniqueness theorem I have long since promised.

4.13. THEOREM (UNIQUENESS OF FIXED POINTS). *Let* (B, A) *be a finite* FPA. *Then every* $\alpha \in B$ *possesses a unique fixed point* $a \in A$.

PROOF. By induction on $|A|$.

For $A = 2$, B consists only of constant functions and the result is true.

Let $|A| > 2$. Suppose, by way of contradiction, that $\alpha \in B$ has two distinct fixed points a, b.

Case 1. $b \neq a'$. Then $|[a \cdot b, a + b]| < |A|$ and we can apply the induction hypothesis to $(B^r, [a \cdot b, a + b])$:

$$a = \alpha a \Rightarrow a = \alpha^r a, \qquad b = \alpha b \Rightarrow b = \alpha^r b,$$

whence α^r has two distinct fixed points, a contradiction.

Case 2. $b = a'$. Since $|A|$ is finite, there is some n such that $\alpha^n \restriction \text{range}(\alpha^n)$ is the identity.

If $\text{range}(\alpha^n) = \{a, b\}$, then $\beta = \lambda x.(\alpha^n x)'$ has no fixed point. Thus, there is some $c \in \text{range}(\alpha^n)$ with $c \neq a, b$. Then $\alpha^n \in B$ has fixed points a, b, and c. Considering a and c, we are in Case 1 relative to α^n—hence in another contradictory state. Q.E.D.

4.14. COROLLARY. *Let* (B, A) *be a finite* FPA.

(i) $\forall \alpha \in B\, \exists n \in \omega [\alpha^n$ *is constant*$]$.

(ii) $\forall \alpha \in B$, α *is not a homomorphism of* A.

I remark that 4.14(ii) shows finite FPA's cannot adequately model the structure $(A_{\text{PA}}; \mu)$ where μ is a definable homomorphism of A_{PA}.

With these applications, we see that the induced algebra constitutes a nice piece of machinery. But it is not enough. We need an additional tool—a group operation. Fortunately, every boolean algebra comes equipped with such—define

$$x \dot{+} y = (x + y) \cdot (x' + y') = x \cdot y' + yx'.$$

(This notation is truly old-fashioned. Modernists would call $\dot{+}$ the symmetric difference and write $x \Delta y$. Since our use of this operation will be group theoretic, it seems more appropriate to use an addition symbol of some kind.)

When we realise we have a group B acting on a group A in an FPA (B, A), we can apply known tools of finite group theory—particularly counting arguments—to our study of FPA's. One such tool for counting is the Burnside Counting Theorem (cf. e.g. [7]) from which the uniqueness theorem immediately follows. I won't present the Burnside Counting Theorem here, but I will derive the uniqueness theorem group theoretically—both to illustrate the technique and to establish a useful lemma.

4.15. DEFINITION. Let (B, A) be an FPA and $a \in A$. Define

$$\mathfrak{F}_a = \{\alpha \in B : \alpha a = a\}.$$

It is probably worth remarking that $\mathfrak{F}_a$ is always a sublattice of B, and $\mathfrak{F}_1$ is a filter and $\mathfrak{F}_0$ an ideal of B. The fact about the sets $\mathfrak{F}_a$ that is of greatest immediate interest, however, is the following:

4.16. LEMMA. *Let* (B, A) *be a finite* FPA *and* $a, b \in A$. *Then* $|\mathfrak{F}_a| = |\mathfrak{F}_b|$ $(= |B|/|A|)$.

PROOF. Define, for each pair $c, d \in A$,

$$\mathfrak{P}_{cd} = \{\alpha \in B: \alpha c = d\}.$$

CLAIM. For each c, d, $|\mathfrak{P}_{cd}| = |\mathfrak{P}_{c0}|$.

To see this, note simply that, for fixed α_0, the map $\alpha \mapsto \alpha \dot{+} \alpha_0$ maps B one-one onto itself and, if $\alpha_0 c = d$ (as does, e.g., the constant function $\lambda x.d$), it maps $\mathfrak{P}_{cd}$ one-one onto $\mathfrak{P}_{c0}$ (since $x \dot{+} x = 0$ for all $x \in A$).

But, for fixed c, B is the disjoint union of $\{\mathfrak{P}_{cd}: d \in A\}$. Thus,

$$|\mathfrak{P}_{cd}| = (1/|A|) \cdot |B| \quad \text{and} \quad |\mathfrak{F}_c| = |\mathfrak{P}_{cc}| = |B|/|A|. \quad \text{Q.E.D.}$$

The uniqueness theorem is an easy corollary to Lemma 4.16.

SECOND PROOF OF THEOREM 4.13. Since every $\alpha \in B$ has a fixed point, $B = \bigcup_{a \in A} \mathfrak{F}_a$. Thus, $|B| \leqslant \sum_{a \in A} |\mathfrak{F}_a|$. Clearly we can have equality only if the $\mathfrak{F}_a$'s are disjoint. But we have equality,

$$\sum_{a \in A} |\mathfrak{F}_a| = \sum_{a \in A} (|B|/|A|) = |A|(|B|/|A|) = |B|. \quad \text{Q.E.D.}$$

I remark that this proof is conceptually simpler than the earlier one and uses a bit less of the structure of FPA's.

Lemma 4.16 has another application—the characterisation of FPA's on 4.

4.17. COROLLARY. *The only algebras* B *for which* $(B, 4)$ *is an* FPA *are those of Examples* 4.2.

PROOF. We first use Lemma 4.16 to reduce this result to the following corollary to uniqueness.

4.18. LEMMA. *Let* $(B, 4)$ *be an* FPA *and* $\alpha \in B$ *such that* $\alpha 1 = 1$. *Then* α *is one of the* τ*-operators*:

x	0	a	a'	1
$\tau_0 x$	1	1	1	1
$\tau_1 x$	a	1	a	1
$\tau_2 x$	a'	a'	1	1

Let $(B, 4)$ be an FPA. By the lemma, the only functions $\alpha \in B$ with $\alpha 1 = 1$ are τ_0, τ_1, and/or τ_2, i.e. $\mathfrak{F}_1 \subseteq \{\tau_0, \tau_1, \tau_2\}$. Now, by Lemma 4.16, $|B| = 4|\mathfrak{F}_1|$. Since $|B|$ is a power of 2, $|\mathfrak{F}_1| \neq 3$ and $\mathfrak{F}_1 \neq \{\tau_0, \tau_1, \tau_2\}$. But τ_0 is in B, whence $\mathfrak{F}_1 = \{\tau_0\}$ or $\{\tau_0, \tau_1\}$ or $\{\tau_0, \tau_2\}$. In the first case $|B| = 4$ and B is the algebra B_0 of constant functions of Example 4.2(i); in the second case $|B| = 8$ and B (contains and hence) is the algebra B_1 of 4.2(ii); and in the third case B is the algebra B_2 of 4.2(iii).

It remains only to prove the lemma.

PROOF OF LEMMA 4.18. Let $\alpha \in B$ with $\alpha 1 = 1$.

Case 1. $\alpha 0 = 0$. Impossible by the uniqueness of fixed points.

Case 2. $\alpha 0 = 1$. Again uniqueness yields $\alpha a \neq a$ and $\alpha a' \neq a'$. The uniqueness of the fixed point of α' similarly yields $\alpha a \neq a'$ and $\alpha a' \neq a$. Thus $\alpha a, \alpha a' \in \{0, 1\}$.

Suppose $\alpha a = 0$. Then $0, a$ are both fixed points of $\lambda x.(a \cdot \alpha' x)$. Thus, $\alpha a = 1$. Similarly, $\alpha a' = 1$. This yields $\alpha = \tau_0$.

Case 3. $\alpha 0 = a$. Now $\alpha a \notin \{0, a\}$ since (by 4.14(i)) $\lim_{n \to \infty} \alpha^n(a) = 1$. Thus $\alpha a \in \{a', 1\}$.

Suppose $\alpha a = a'$. By 4.14(i), $\alpha a' \notin \{0, a, a'\}$. Thus, α has the table

x	0	a	a'	1
αx	a	a'	1	1

But $\beta = \lambda x.(a' \cdot \alpha x)$ has fixed points 0 and a', a contradiction. Thus $\alpha a = 1$.

Now α' has the partial table

x	0	a	a'	1
$\alpha' x$	a'	0	?	0

For α' to have a fixed point, we must have $\alpha' a' = a'$, i.e. $\alpha a' = a$. The function just determined is $\alpha = \tau_1$.

Case 4. $\alpha 0 = a'$. By a symmetric argument, $\alpha = \tau_2$. Q.E.D.

The final result of this section is really out of place here—it properly belongs in the beginning of the next section. I put it here, however, because it concerns finite FPA's in general and the next section is devoted to finite *closed* FPA's. By way of motivation, I first digress to note the following:

4.19. LEMMA. *Let* (B, A) *be a finite diagonalisable* FPA *and let* τ *be any* τ*-operator such that* B *consists of polynomials in* τ. *Then: If an atom* $a \in A$ *is minimal in the dual ordering* $\prec$ *on the atoms of* A, *then, for all* $x \in A$ *and all* $\alpha \in B$, $\alpha x = \alpha(x + a)$.

This fact is quickly discovered empirically. A proof is by induction on the generation of α as a polynomial in τ and is left to the reader as an exercise. The result is important insofar as it gives us a special choice of atom a to which to apply the induced algebra apparatus: If a satisfies the conclusion of Lemma 4.19, the induced algebra $(B', [a, 1])$ preserves an optimal amount of the structure of (B, A). We will see this in the next section. For the moment, our goal is to come as close to constructing such atoms as possible without assuming any additional closure condition on an FPA. In this endeavour, I have not been very successful and the result (Lemma 4.21) is quite weak.

4.20. DEFINITION. Let (B, A) be an FPA. Define an equivalence relation $\sim$ on A by: for $a, b \in A$

$$a \sim b: \forall \alpha \in B(\alpha a = \alpha b).$$

4.21. LEMMA. *Let* (B, A) *be a finite* FPA. *For each* $a \in A$ *there is an element* $b \in A$ *with* $a \neq b$, *but* $a \sim b$; *i.e. the* $\sim$*-equivalence classes all have cardinality at least* 2.

PROOF. By contradiction. Suppose $a \in A$ is such that, for each $b \in A$, there is an $\alpha_b \in B$ with $\alpha_b a \neq \alpha_b b$.

CLAIM. We can assume $\alpha_b a = 0$, $\alpha_b b > 0$.

To prove the claim let $\alpha_b a = c$, $\alpha_b b = d$, and consider the various cases:

$c < d$. Let $\beta = \lambda x.(c' \cdot \alpha_b x)$. Then $\beta a = 0$, $\beta b > 0$.

$c > d$. Let $\beta = \lambda x.(c \cdot \alpha_b' x)$.

$c \| d$, i.e. c and d are incomparable. Again define $\beta = \lambda x.(c' \cdot \alpha_b x)$.

This proves the claim.

Now define $\beta = \lambda x.\Sigma\{\alpha x\colon\ \alpha \in B \,\&\, \alpha a = 0\}$, where Σ, being undotted, denotes the supremum. Then $\beta a = 0$ and, for all $b \neq a$, $\beta b \neq 0$. Now let

$$\gamma = \lambda x.\,(a \dot{+} \beta x).$$

The function γ has the peculiar property of mapping a to a and nothing else to a—which property conflicts with Corollary 4.14(i) whereby $\lim_{n\to\infty} \gamma^n b = a$ for all $b \in A$. Q.E.D.

Our most important application of Lemma 4.21 will be given in the next section. Before turning to this, let me quickly exploit it to improve some bounds.

4.22. COROLLARY. *Let* (B, A) *be an* FPA *with* $|A| = 2^n$. *Then* $|B| \leqslant |A| \cdot (|A| - 2)^{2|A|/2-1}$.

PROOF. A decomposes into a disjoint union of $\sim$-equivalence classes. Since each class has at least 2 elements, there are at most $\frac{1}{2}|A|$ such classes. Now, how many $\alpha \in B$ can there be?

Case 1. There are $\frac{1}{2}|A|$ classes. Then each class consists of exactly 2 elements and $\alpha \in B$ is determined by (i) choosing one of $\frac{1}{2}|A|$ classes to contain the fixed point, (ii) choosing one of 2 elements of this class to be the fixed point, and (iii) choosing, for each of the $\frac{1}{2}|A| - 1$ other classes, one of $|A| - 2$ elements outside the class to map the class onto. Put together, these factors yield the stated bound.

Case 2. There are fewer than $\frac{1}{2}|A|$ classes. Then, corresponding to the list above, one has (i) a factor less than $\frac{1}{2}|A|$, (ii) a factor less than or equal to $|A|$, so less than $2(|A| - 2)$ (ignoring trivial A), and (iii) at most $\frac{1}{2}|A| - 2$ factors less than or equal to $|A| - 2$. Q.E.D.

For $A = 8$, this yields

4.23. FIFTH BOUND. Let $(B, 8)$ be an FPA. Then $|B| \leqslant 1024$.

In the next section, we will combine application of Lemma 4.21 and the induced algebra construction to, among other things, reduce this bound to 128 —for algebras $(B, 8)$ which are *closed*.

5. Finite closed fixed point algebras. Let me straightaway define what I mean by a closed FPA.

5.1. DEFINITION. An FPA (B, A) is *closed* if, in addition to satisfying conditions (i)–(v) of Definition 1.10, (B, A) satisfies the closure condition:

(vi) for all $a \in A$ and $\alpha \in B$, $\lambda x.\alpha(x \dot{+} a) \in B$ and $\lambda x.\alpha(a \cdot x) \in B$.

Note that condition (vi) is equivalent to

(vi′) for all $\alpha \in B$ and any boolean polynomial $p(x)$ over A, $\lambda x.\alpha\, p(x) \in B$.

The choice of (vi) over (vi′) is both a matter of economy and heuristic—the latter since it is the form in which we apply the closure.

We have already seen examples of FPA's that are not closed—cf. the third incarnation of Examples 4.3—and shall shortly see another (5.5, below). What about examples of closed FPA's? It happens that the logically interesting ones—the Lindenbaum FPA's—are closed. So, too, are the diagonalisable ones. Indeed, from a logical point of view, since they more closely model the Lindenbaum FPA's, closed FPA's are more central objects than FPA's in general. Be that as it may, our present interest in them is in that the finite closed FPA's have the special atoms found, according to Lemma 4.19, in finite DA's.

5.2. DEFINITION. Let (B, A) be an FPA. An atom $a \in A$ is a *fundamental atom* for (B, A) if $\forall x \in A \forall \alpha \in B(\alpha x = \alpha(x + a))$.

In terms of the relation $\sim$ of Definition 4.20, an atom $a \in A$ is fundamental if $\forall x \in A(x \sim x + a)$. Note that this is equivalent to $\forall x \in A(x \sim x \dot{+} a)$.

5.3. THEOREM. *Let* (B, A) *be a finite closed* FPA. *Then* (B, A) *possesses a fundamental atom.*

The proof is given via two simple lemmas.

5.4. LEMMA. *Let* (B, A) *be an* FPA *closed under*

$$\alpha \in B \Rightarrow \lambda x.\alpha(x \dot{+} c) \in B$$

for all $c \in A$. *Then*: *For any* $a, b \in A$,

$$a \sim b \Leftrightarrow a \dot{+} b \sim 0.$$

PROOF. Suppose $a \sim b$. Let $\alpha \in B$ be given and $\beta = \lambda x.\alpha(x \dot{+} b)$. Now

$$a \sim b \Rightarrow \beta a = \beta b \Rightarrow \alpha(a \dot{+} b) = \alpha(b \dot{+} b) \Rightarrow \alpha(a \dot{+} b) = \alpha 0.$$

Conversely, suppose $a \dot{+} b \sim 0$. Let α be given and again define $\beta = \lambda x.\alpha(x \dot{+} b)$. As before

$$a \dot{+} b \sim 0 \Rightarrow \beta(a \dot{+} b) = \beta 0$$

$$\Rightarrow \alpha(a \dot{+} b \dot{+} b) = \alpha(0 \dot{+} b) \Rightarrow \alpha a = \alpha b. \quad \text{Q.E.D.}$$

Let me quickly remark that the stated closure condition is necessary for the general truth of the lemma.

5.5. COUNTEREXAMPLE. Let $A = 8$ and, for the labelling of the elements of 8 of the beginning of the last section, let $\alpha: A \to A$ be defined by

x	$0, c$	$a, b, d, e, f, 1$
αx	1	d

Then, in $(A[\alpha], A)$ (which is a subdiagonalisable FPA), we have $a \sim b \sim d \sim e \sim f \sim 1$ and $0 \sim c$, but $a \dot{+} b = d \not\sim 0$.

Continuing our proof that finite closed FPA's have fundamental atoms, we now see how the second new closure property is used.

5.6. LEMMA. *Let* (B, A) *be an* FPA *closed under*

$$\alpha \in B \Rightarrow \lambda x.\alpha(c \cdot x) \in B$$

for all $c \in A$. *Then: For any* $a, b \in A$,

$$a \sim 0 \Rightarrow a \cdot b \sim 0.$$

PROOF. Suppose $a \sim 0$ and $\alpha \in B$. Let $\beta = \lambda x.\alpha(b \cdot x)$. Note

$$a \sim 0 \Rightarrow \beta a = \beta 0 \Rightarrow \alpha(a \cdot b) = \alpha(0 \cdot b) \Rightarrow \alpha(a \cdot b) = \alpha 0. \quad \text{Q.E.D.}$$

PROOF OF THEOREM 5.3. Let (B, A) be a finite closed FPA. By Lemma 4.21, there is an element $a \in A$ such that $a \sim 0$. Let $b \leqslant a$ be any atom. By Lemma 5.6, $b \sim 0$. By Lemma 5.4, for any $x \in A$, $x \dot{+} b \sim x$. Q.E.D.

I said on introducing them that our present interest in finite closed FPA's lay in their possession of fundamental atoms. Basically this is true, but we need an hereditary possession of fundamental atoms. Recall our first proof, in the last section, of the uniqueness of fixed points. This was an induction involving the induced algebra. To carry out such an induction for a subclass of finite FPA's, we must know the subclass to be closed under the operation of inducing the algebras. The subclass of finite FPA's possessing fundamental atoms does not have this closure property unless all finite FPA's possess fundamental atoms. But the class of closed FPA's does.

5.7. LEMMA. *Let* (B, A) *be a closed* FPA, $a, b \in A$ *with* $a < b$. *Then the induced algebra* $(B^r, [a, b])$ *is closed.*

The proof is trivial.

Lemma 5.7 allows us to use the induced algebra construction to give an inductive proof of the following:

5.8. THEOREM (FUNDAMENTAL THEOREM). *Let* (B, A) *be a finite closed* FPA. *Then* (B, A) *is subdiagonalisable.*

The assertion of subdiagonalisability means, again, that there is a τ-operator τ on A such that every $\alpha \in B$ is a polynomial in τ of the proper sort over A.

Simultaneously with Theorem 5.8 we prove the following:

5.9. THEOREM. *Let* (B, A) *be a closed* FPA *with* $|A| = 2^n$. *Then*

$$|B| \leqslant 2^{2^n - 1} = 2^{|A|-1}.$$

Moreover, equality holds in exactly $n!$ *cases.*

Theorems 5.8 and 5.9 are proven simultaneously by induction on $|A| = 2^n$, i.e. by induction on n. The case $n = 1$ is trivial and the case $n = 2$ was proven in the last section. For the case $n = 3$, i.e. $A = 8$, let us conclude from 5.9 our final bound.

5.10. SIXTH BOUND. Let $(B, 8)$ be a closed FPA. Then $|B| \leqslant 128$.

We have already encountered a DA on 8 of this cardinality and so can conclude the bound to be best possible. [That is, it is best possible numerically. For $A = 8$, we do not need to assume the closedness of $(B, 8)$ to conclude Lemma 5.7, whence the succeeding proofs, when specialised to $A = 8$, yield the

Sixth Bound under the weaker assumption that $(B, 8)$ possesses a fundamental atom.]

There are two steps to proving 5.8 and 5.9. First, we let (B, A) be a finite closed FPA with A of cardinality 2^{n+1} and show B to be a subset of one of at most $(n+1)!$ boolean algebras B_0 such that (B_0, A) is a closed FPA with $|B_0| = 2^{2^{n+1}-1}$. Second, we exhibit $(n+1)!$ distinct such algebras (B_0, A)—each of which is diagonalisable. When these two steps are completed, the theorems immediately follow.

The first step is fairly simple. First we prove a little lemma.

5.11. LEMMA. *Let* (B, A) *be a finite closed* FPA *with a fundamental atom* $a \in A$. *Let* $(B', [a, 1])$ *be the induced algebra and* $C \supseteq B'$ *an algebra such that* $(C, [a, 1])$ *is a closed* FPA. *Define*

$$B_0 = \{\alpha : A \to A : \alpha^r \in C \,\&\, \forall x \in A[\alpha x = \alpha(x + a)]\}.$$

Then

(i) $B \subseteq B_0$,

(ii) (B_0, A) *is a closed* FPA *with fundamental atom* a,

(iii) $|B| \leqslant |B_0| = |C| \cdot 2^{|A|/2}$.

PROOF. Note first that we can define α^r for any $\alpha : A \to A$, not merely for $\alpha \in B$. Note too that, by Lemma 5.7, some such C exists. Thus, the definition of B_0 makes sense.

Assertion (i) is immediate. Assertion (iii) is a direct calculation: There are $|C|$ choices of $\alpha^r \in C$ (the first factor). For each such α^r, there are $2^{|A|/2}$ possible $\beta \in B_0$ (the second factor) for which $\beta^r = \alpha^r$. For: (i) for any given value $\alpha^r x$, there are two possible values for βx—namely, $\alpha^r x$ and $a \dot{+} \alpha^r x$; and (ii) by the assumption that $\alpha x = \alpha(x + a)$, this choice need only be made once for each pair x, $x \dot{+} a$, and there are $|A|/2$ such pairs.

(ii) It remains only to see that (B_0, A) is a closed FPA with fundamental atom a. Given the rest, the fact that a is a fundamental atom is trivial.

B_0 is a boolean algebra. Let $\alpha, \beta \in B_0$ and let $\gamma = \alpha + \beta$. Note that

$$\alpha^r \in C \,\&\, \beta^r \in C \Rightarrow (\alpha + \beta)^r = \alpha^r + \beta^r \in C,$$

$$\forall x \in A[\alpha x = \alpha(x + a) \,\&\, \beta x = \beta(x + a)] \Rightarrow \forall x \in A[\gamma x = \gamma(x + a)].$$

Hence $\gamma \in B_0$. Similarly, if $\alpha \in B_0$, then $\alpha' \in B_0$.

B_0 is closed under composition. Let $\alpha, \beta \in B_0$, $\gamma = \lambda x.\alpha\beta x$. Now, for any $x \in A$,

$$\begin{aligned} \gamma^r x &= a + \alpha\beta x = a + \alpha(a + \beta x), \quad \text{since } \forall y \in A(\alpha y = \alpha(y + a)) \\ &= \alpha^r \beta^r x, \end{aligned}$$

and we know C is closed under composition. Thus $\gamma^r \in C$. But also, for any $x \in A$,

$$\gamma x = \alpha\beta x = \alpha\beta(x + a) = \gamma(x + a),$$

since $\beta x = \beta(x + a)$. Thus $\gamma \in B_0$.

B_0 has fixed points. Let $\alpha \in B_0$. Suppose $b \in [a, 1]$ is a fixed point of $\alpha^r \in C$. Now either $\alpha b = \alpha^r b = b$, and b is a fixed point of α, or

$$\alpha(a \dot{+} b) = \alpha b = a \dot{+} \alpha^r b = a \dot{+} b,$$

and $a \dot{+} b$ is a fixed point.

It follows that (B_0, A) is an FPA. All that remains is the following:

(B_0, A) is a closed FPA. Let $\alpha \in B_0$, $b \in A$, and consider $\beta = \lambda x.\alpha(x \dot{+} b)$. A tiny argument shows that, for all $x \in A$, $\alpha(x \dot{+} b) = \alpha(x \dot{+} (b + a))$, whence we can assume without loss of generality $b \in [a, 1]$. Since C is closed, $\beta^r = \lambda x.\alpha^r(x \dot{+} b) \in C$. But, for any $x \in A$,

$$\beta x = \alpha(x \dot{+} b) = \alpha((x \dot{+} b) \dot{+} a) = \alpha((x \dot{+} a) \dot{+} b) = \beta(x \dot{+} a).$$

Since $x + a = x$ or $x + a = x \dot{+} a$, it follows that $\beta x = \beta(x + a)$ and $\beta \in B_0$. Similarly, $\gamma = \lambda x.\alpha(b \cdot x) \in B_0$. Q.E.D.

Now, what about our cardinal calculation? Suppose 5.9 holds for n and $|A| = 2^{n+1}$. Then

$$|B_0| = |C| \cdot 2^{2^n} \leqslant 2^{2^{n-1}} \cdot 2^{2^n} = 2^{2^{n+1}-1},$$

with equality only if $|C| = 2^{2^{n-1}}$. By induction hypothesis, B^r is contained in some C of cardinality $2^{2^{n-1}}$, whence B is contained in some B_0 of cardinality $2^{2^{n+1}-1}$.

As for the number of FPA's on A with $|B| = 2^{2^{n+1}-1}$, note that there are $n + 1$ choices of a fundamental atom a and, given this choice, there are, by induction hypothesis, $n!$ choices of FPA's $(C, [a, 1])$, with $|C| = 2^{2^{n-1}}$. Thus, there are at most $(n + 1) \cdot n! = (n + 1)!$ FPA's of maximum cardinality on A.

Note that I wrote "at most". We must verify that no two of the $(n + 1)!$ algebras of maximum cardinality coincide. This is precisely the second step of the proof described above—for a given k, we find $k!$ distinct DA's (B, A) on an algebra A of cardinality 2^k with $|B| = 2^{2^{k-1}}$. Letting $k = n + 1$ will, of course, complete the induction step in the proof of Theorem 5.9. It will also yield Theorem 5.8 since all FPA's of maximum cardinality are diagonalisable.

5.12. LEMMA. *Let $k \geqslant 2$ be given and A a boolean algebra of cardinality 2^k with set of atoms $P = \{a_0, \ldots, a_{k-1}\}$. Given a total ordering $\prec$ of P, let (B, A) be the corresponding diagonalisable* FPA *(i.e. the* FPA *in which B consists of all appropriate polynomials in the operator τ dual to $\prec$). Then*

(i) *the $k!$ distinct total orderings of P yield distinct* FPA*'s*

(ii) $|B| = 2^{2^{k-1}}$.

PROOF. (i) Let $\prec_1$, $\prec_2$ be distinct total orderings of P and let τ_1, τ_2 be their respective τ-operators. We will see that τ_1, τ_2 are incompatible, i.e. that they generate an $\alpha: A \to A$ with two fixed points. In fact, we can take $\alpha = \tau_1 \cdot \tau_2$.

Suppose $\prec_1$ and $\prec_2$ look like

$$a_{i_{k-1}} \prec_1 \cdots \prec_1 a_{i_0}, \qquad a_{j_{k-1}} \prec_2 \cdots \prec_2 a_{j_0},$$

and m is minimum such that $a_{i_m} \neq a_{j_m}$. Let $b = a_{i_0} + \cdots + a_{i_{m-1}} = a_{j_0} + \cdots + a_{j_{m-1}}$ ($= 0$ if $m = 0$). Now $\tau_1 b = b + a_{i_m}$ and $\tau_2 b = b + a_{j_m}$, whence

$$\tau_1 b \cdot \tau_2 b = (b + a_{i_m})(b + a_{j_m}) = b,$$

since a_{i_m}, a_{j_m} are distinct atoms incomparable to b. But we also have $\tau_1 1 \cdot \tau_2 1 = 1 \cdot 1 = 1$.

(ii) By induction on k. For $k = 2$, cf. Examples 4.2.

Let (B, A) be the diagonalisable FPA generated by a linear ordering $\prec$ of P with the atom a minimum in $\prec$. Consider the induced algebra $(B^r, [a, 1])$. Since a is a fundamental atom (by Lemma 4.19) Lemma 5.11 applies and

$$|B_0| = |B^r| \cdot 2^{2^{k-1}},$$

where B_0 is the maximum algebra inducing B^r.

We first observe that $B = B_0$: Let $\alpha \in B_0$. There is an $\alpha_0 \in B$ such that $\alpha_0^r = \alpha^r$. Let $\alpha_1 = \lambda x.(a' \cdot \alpha_0 x) \in B$. α_1 is the minimum function inducing α^r. Note: For any $b \in A$,

$$a \cdot \tau b = \begin{cases} a, & b = 1 \text{ or } b = 1 \dot{+} a \qquad (\text{i.e. } b \sim 1), \\ 0, & \text{otherwise.} \end{cases}$$

Let $b_0, \ldots, b_m \in A$ enumerate all $b \in A$ such that $\alpha b = \alpha_1 b + a$. Then

$$\alpha = \lambda x.\left[\alpha_1 x + a \sum_i \tau(x \dot{+} b \dot{+} 1)\right] \in B.$$

Thus, $|B| = |B^r| \cdot 2^{2^{k-1}}$, and it suffices to show that $(B^r, [a, 1])$ is (up to labelling—i.e. isomorphism) the diagonalisable algebra on the ordering of $P - \{a\}$ induced by $\prec$. For, then

$$\begin{aligned} |B| &= 2^{2^{k-1}-1} \cdot 2^{2^{k-1}}, \quad \text{by induction hypothesis,} \\ &= 2^{2^{k-1}}. \end{aligned}$$

First note that $\tau^r = \lambda x.(a + \tau x)$ is a τ-operator on $[a, 1]$: For all x, $y \in [a, 1]$,

(1) $\tau^r 1 = a + \tau 1 = a + 1 = 1$.

(2) $\tau^r x \cdot \tau^r(x \to y) = (a + \tau x)(a + \tau(x \to y)) = a + \tau x \cdot \tau(x \to y) \leqslant a + \tau y = \tau^r y$.

(3) $\tau^r x = a + \tau x \leqslant a + \tau\tau x = a + \tau(a + \tau x) = \tau^r \tau^r x$.

(4) $\tau^r x \leqslant x \Rightarrow a + \tau x \leqslant x \Rightarrow \tau x \leqslant x \Rightarrow x = 1$.

[Note: Given the existence of fixed points, this last verification was unnecessary.]

Second, note that the atoms of $[a, 1]$ are given by $P^r = \{a + a_i : a \neq a_i \in P\}$, and so are in obvious correspondence with $P - \{a\}$. What is the ordering $\prec^r$ on P^r dual to τ^r? Note, for $a_i, a_j \in P - \{a\}$,

$$\begin{aligned} & a_i \prec a_j \Rightarrow \forall x \in A[a_i \leqslant \tau x \Rightarrow a_j \leqslant x] \\ (*) \qquad & \Rightarrow \forall x \in A[a + a_i \leqslant a + \tau x \Rightarrow a + a_j \leqslant a + x] \\ & \Rightarrow \forall x \in [a, 1][a + a_i \leqslant \tau^r x \Rightarrow a + a_j \leqslant x] \\ & \Rightarrow a + a_i \prec^r a + a_j, \end{aligned}$$

where (*) follows since a is an atom disjoint from a_i, a_j, and τx—unless $\tau x = 1$. Since $\prec$ is total, so too must $\prec^r$ be.

Third, we must see that B^r includes all appropriate polynomials in τ^r. But this follows quickly from the facts that B^r is closed under boolean operations, compositions, and (by Lemma 5.7) the abstractions

$$\alpha^r \in B^r \Rightarrow \lambda x.\alpha^r(x \dot{+} b), \lambda x.\alpha^r(b \cdot x) \in B^r,$$

for all $b \in [a, 1]$.

Fourth, and last, we must note that B^r contains nothing more than the appropriate polynomials in τ^r. There are two possibilities for this proof. We can appeal to the induction hypothesis of Theorem 5.9—$|B^r| = 2^{2^{k-1}-1}$ and so B^r cannot be further extended (letting $k = n + 1$); or, we can simply note that the generation of α from τ in B induces such a generation of α^r from τ^r in B^r. Q.E.D.

As I noted earlier, with the proof of this lemma completed, we have completed the proofs of Theorems 5.8 and 5.9.

Algebraically, there is yet much room for improvement. Aside from needing the closedness of (B, A) to yield hereditarily the existence of fundamental atoms, we only needed this closedness to conclude the closedness of the FPA's we constructed. Hence all algebras (B, A) hereditarily possessing fundamental atoms are subdiagonalisable and, should all FPA's possess fundamental atoms, all FPA's will be subdiagonalisable. Also, it follows that, more efficiently applied, the closedness of a finite closed FPA should yield a stronger conclusion, e.g. that the given closed FPA is, in fact, diagonalisable.

But, if the result can be improved algebraically, it is not in much need of logical improvement: It is sufficiently powerful to tell us that the marvellous success reported on in §3 in analysing arithmetic via finite algebras cannot be much improved—only those operators $\alpha \in B_{\mathrm{PA}}$ which somehow reduce to τ-operators can be given an adequate finite algebraic modelling.

6. Infinite fixed point algebras. We have seen that, logically speaking, finite FPA's offer no great advance over finite DA's. However, this does not mean we should dismiss FPA's as a logical loss—there are still the infinite FPA's to consider. The finite and infinite FPA's are vastly different. If the former are inadequate for modelling extensional arithmetic self-reference, the latter certainly are not. Nonetheless, I make no claims of arithmetic applicability of the infinite algebraic modelling. The rôle played here by the infinite algebra is descriptive rather than instrumental. But this is standard in logic, where many a lattice and semilattice has been studied by purely nonalgebraic techniques. However, as FPA's may be of some independent algebraic interest, I will first briefly discuss some nonarithmetic matters.

The first thing to do is to formally note the difference between the finite and infinite FPA's.

6.1. REMARKS. (i) The classes of finite and of all FPA's do not even share a common universal theory. The finite FPA's all satisfy, e.g., the uniqueness assertion,

$$\forall \alpha \in B \forall x, y \in A[\alpha x = x \wedge \alpha y = y \supset x = y].$$

[One could ask about the equational theories—but, with no function constants other than $\lambda x.0$ and $\lambda x.1$, these are likely to be pretty trivial theories.]

(ii) Both the first-order theories of the finite and of all the FPA's are undecidable. [In fact, the proof of finite inseparability for DA's in [**49**] adapts quite easily to FPA's.]

The next best thing to a finite FPA is an infinite DA. These can be produced either via the representation theory for DA's or via the Lindenbaumisation process. Let me first discuss the former. Between the rather restrictive requirement of *relative* reverse well-foundedness (most easily satisfied in smaller algebras A) and the more inclusive one of *booleanness* (most easily satisfied in larger algebras), the representation theory is not easy to apply directly. A simple variant of this representation theory, however, makes possible the easy exhibition of many examples.

6.2. LEMMA. *Let X_0 be a set, $R = \prec$ a reverse well-founded partial ordering of X_0, and $A \subseteq P(X_0)$ any algebra satisfying R^{-1}: $A \to A$ (e.g. $A = P(X_0)$). Then the function τ: $A \to A$ defined by*

$$\tau x = \{a \in X_0 : \forall b \in X_0 (a \prec b \Rightarrow b \in x)\} = X_0 - R^{-1}(X_0 - x)$$

is a τ-operator making $(A; \tau)$ a DA.

Note that this is *not* a special case of Theorem 3.6 as X_0 need not be the set X of *all* ultrafilters of A. The proof, however, is much the same and need not be presented here. In any event, it can be found in any number of places in the literature.

6.3. EXAMPLES (EXERCISES). (i) Let $X_0 = \{a_n : n \in \omega\}$, $X_1 = X_0 \cup \{a_\omega\}$ be two infinite sets of atoms and let $\prec_0$ and $\prec_1$ order these sets in the types ω^* and $(\omega + 1)^*$, respectively, i.e.

$$\cdots \prec_0 a_2 \prec_0 a_1 \prec_0 a_0,$$
$$a_\omega \prec_1 \cdots \prec_1 a_2 \prec_1 a_1 \prec_1 a_0.$$

Let A_0 be the algebra of finite and cofinite subsets of X_0; and let A_1 be the algebra of finite subsets of X_0 and cofinite subsets of X_1 containing a_ω. Then, letting $R_i = \prec_i$, we have R_i^{-1}: $A_i \to A_i$ and we have τ-operators τ_i dual to $\prec_i$ on A_i. On each algebra A_i, τ_i satisfies: For $x \in A_i$,

$$\tau_i x = \begin{cases} 1, & x = 1, \\ \{a_0, \ldots, a_m\}, & m \text{ is minimum such that } a_m \notin x. \end{cases}$$

Moreover $(A_0; \tau_0)$ is isomorphic to $(A_1; \tau_1)$.

(ii) Let $n > 0$. Let X_0 be as in (i), $Y_n = \{b_0, \ldots, b_{n-1}\}$, and $Y_\omega = \{b_n : n \in \omega\}$. Define A_n^* to be the finite-cofinite algebra on $X_0 \cup Y_n$ for $0 < n \leqslant \omega$ with $\prec_n^*$ on $X_0 \cup Y_n$ extending $\prec_0$ on X_0 by defining $a_m \prec_n^* b_k$ for all $m > k > n$. Then, if $R_n^* = \prec_n^*$, we have $(R_n^*)^{-1}$: $A_n^* \to A_n^*$ and, for τ_n^* the dual operator:

(a) Each $(A_n^*; \tau_n^*)$ is a DA.

(b) No $(A_n^*; \tau_n^*)$ is isomorphic to $(A_0; \tau_0)$.

(c) For $m < n \leqslant \omega$, $(A_m^*; \tau_m^*)$ is not isomorphic to $(A_n^*; \tau_n^*)$.

[Hint for (b), (c). Compare $\tau_0 0$, $\tau_m^* 0$, and $\tau_n^* 0$. The ambitious reader may wish to show that in fact these algebras have distinct equational theories.]

Such examples can be multiplied indefinitely and are generally of limited interest. One exception is the algebra $(A_0; \tau_0)$ of 6.3(i). This is the free DA on 0 generators.

Free diagonalisable algebras have been studied a bit—cf. [**3**, **21**, **27**, **29** and **50**]. They even have some bearing on arithmetic—most notably:

(i) In [**3**], Bernardi made good use of the freedom of $(A_0; \tau_0)$ in anticipation of the de Jongh-Sambin Theorem (2.6, above); and

(ii) Artyomov [**1**], Montagna [**29**], and Visser (unpublished) have shown the free algebra on ω generators to be embeddable in $(A_{\mathrm{PA}}; \tau)$.

[I suppose I should mention that (i) my independent proof (in [**43**]) of Bernardi's result used the finite representation theory and yielded a more effective result; and (ii) none of the proofs of the embeddability result is algebraic—essentially, one combines the effectiveness of Solovay's proof of his Completeness Theorem (3.10) with an additional (nonextensional) self-reference to uniformise Solovay's result and mention, as it were, the embedding as an afterthought.]

As I said, the second easiest infinite DA's to exhibit are the Lindenbaum ones. And here there is some variety (in the nonalgebraic sense of the word). If T is any formal theory containing a minimal amount of arithmetic, then, by Rosser's Theorem [**34**], the Lindenbaum sentence algebra A_T is atomless—hence countable and atomless, i.e. free on ω generators (as a boolean algebra). Thus, A_{PA}, A_{ZF}, and A_{GB} are all isomorphic. But does this isomorphism still hold if we add the τ-operators $\tau[\varphi] = [\mathrm{Pr}_T(\ulcorner \varphi \urcorner)]$ for $T = \mathrm{PA}$, ZF, GB, etc.? Is $(A_{\mathrm{ZF}}; \tau)$ isomorphic to $(A_{\mathrm{ZF}}; \sigma)$, where σ is as in §1? By variants of Solovay's Completeness Theorems (3.10 and 3.11), these structures have, not only a common equational theory, but also a common universal one. There is no particular reason to believe that this agreement extends—nor to believe the opposite. However, there are theories T, with canonical representations $\mathrm{Pr}_T(\cdot)$ of provability, such that the corresponding Lindenbaum algebras $(A_T; \tau)$ are quite definitely distinct from, say, $(A_{\mathrm{PA}}; \tau)$:

(i) Let $T_0 = \mathrm{PA} + \neg\mathrm{Con}_{\mathrm{PA}}$. It is easily seen that $T_0 \vdash \neg\mathrm{Con}_{T_0}$, whence in $(A_{T_0}; \tau)$ one has $\tau 0 = 1$.

(ii) If $T_1 \not\vdash \neg\mathrm{Con}^n_{T_1}$ for any n, i.e. $T_1 \not\vdash \mathrm{Pr}_{T_1}(\ulcorner \mathrm{Pr}_{T_1}(\ulcorner \ldots \urcorner 0 = 1 \ulcorner \ldots \urcorner) \urcorner)$ for any finite nesting of the formula $\mathrm{Pr}_{T_1}(\cdot)$, then (by [**57**]) $(A_{T_1}; \tau)$ has the same equational theory as $(A_{\mathrm{PA}}; \tau)$. The universal theories of $(A_{T_1}; \tau)$ and $(A_{\mathrm{PA}}; \tau)$ can, however, still disagree.

The situation changes dramatically when we consider the full Lindenbaum FPA's of these theories. (Recall Definition 1.9.) In (B_T, A_T), the τ-operator determined by $\mathrm{Pr}_T(\cdot)$ is not distinguished. One can express within the language of FPA's what it means to be a τ-operator, e.g., by the formula

$$\mathrm{Tau}(\alpha)\colon \forall xy \in A\big[\alpha x \leqslant \alpha\alpha x \wedge \alpha x \cdot \alpha(x \to y) \leqslant \alpha y \wedge \alpha 1 = 1\big];$$

but for many theories T there is a dazzling assortment of τ-operators in (B_T, A_T) (cf. [**6**, **14**, **16**, and **33**] for such an assortment of *well-behaved* τ-operators) and no evident way to distinguish the τ-operator we consider canonically associated with T. In particular, our easy distinction between $(A_{\mathrm{PA}}; \tau)$ and $(A_{T_0}; \tau)$, for $T_0 = \mathrm{PA} + \neg\mathrm{Con}_{\mathrm{PA}}$ is no longer easy and perhaps

not even possible: B_{PA} contains the trivial τ-operator $\lambda x.1$ which maps 0 to 1; and B_{T_0} contains τ-operators which do not map 0 to 1. Let me thus be emboldened to make a conjecture.

6.4. CONJECTURE. Let $T_0 = PA + \neg Con_{PA}$. The Lindenbaum FPA's (B_{PA}, A_{PA}) and (B_{T_0}, A_{T_0}) are isomorphic.

A safer conjecture seems to be

6.5. CONJECTURE. The Lindenbaum FPA's (B_{PA}, A_{PA}) and (B_{ZF}, A_{ZF}) are isomorphic.

The reason for this conjecture is empirical: All known applications of self-reference to the study of PA carry over to that of ZF, and conversely. [Of course, it should be noted that there has been no discernible attempt to apply self-reference to specifically set theoretic formulae, and hence that the empirical evidence is superficial.]

I have one more conjecture—one that might surprise a few readers.

6.6. CONJECTURE. The Lindenbaum FPA's (B_{ZF}, A_{ZF}) and (B_{GB}, A_{GB}) are *not* isomorphic; and, in fact, these structures have distinct first-order theories.

Why should I expect such disparate types of theories as PA and $PA + \neg Con_{PA}$ to have identical self-referential structures and such closely related theories as ZF and GB to have distinct such structures? Well, as I already noted, the easiest distinction between PA and $PA + \neg Con_{PA}$ vanishes when one looks at all extensional formulae. Moreover, many results initially provable only for sound theories like PA (e.g. incompleteness via Gödel's self-referential sentence) extend fairly uniformly to unsound theories like $PA + \neg Con_{PA}$ (here: incompleteness via Rosser's self-referential sentence). While not every such result extends (cf. e.g. [**44**] for some counterexamples), every one expressible in the language of FPA's that I know of has a proof valid for PA and $PA + \neg Con_{PA}$. In other words, every sentence I know true about (B_{PA}, A_{PA}) I also know true about (B_{T_0}, A_{T_0}), for $T_0 = PA + \neg Con_{PA}$; moreover, I know both truths for a *common* reason. This is not the case with ZF and GB: For $T = ZF$ or GB, there is an element $\mu \in B_T$ such that $\mu: A_T \to A_T$ homomorphically. The original constructions of the functions μ were different in the two cases—for ZF one applies the Orey Compactness Theorem (for which see [**6** and **32**]), while for GB it is an interesting construction due to Solovay (cf. [**51**]). There was no common construction and the rôles traditionally played by the homomorphisms μ in GB and ZF are different—as first realised by Hájek in [**12**], and since refined in [**10**] (cf. [**45** or **46**] for an explanation), [**13**, **51**, and **54**]. Of course, Conjecture 6.6 is a conjecture because this difference has yet to be expressed in terms of the FPA's.

[NOTE ADDED IN PROOF. The preceding remarks are insipid at best. Originally I had overestimated and, thus, strongly overemphasised the role of the homomorphisms μ. The original constructions referred to were given by relative interpretations. The converse, which I accepted all too hastily, that the homomorphisms give rise to relative interpretations is simply false—as was pointed out to me in a letter I received from Petr Hájek as I was correcting the galleys.

At the same time, I received from Robert Solovay an affirmative solution to Conjectures 6.4 and 6.5. I have not had the opportunity to study his proof and

see where it fails to refute Conjecture 6.6; but the nature of his solution leads me to believe that the *full* Lindenbaum FPA's offer only an incomplete explication of the extensional self-referential structure of theories.]

B. Bibliography. The following bibliography is lightly padded. It includes, of course, all items referred to in the text, as well as my Handbook article [**41**] for logical background, several papers on self-reference, and numerous uncited references on diagonalisable algebras. The reader interested in self-reference is particularly referred to [**44**, **46**], and their bibliographies.

BIBLIOGRAPHY

1. S. N. Artyomov, *Arithmetically complete modal theories*, Semiotics and Information Science, no. 14, Akad. Nauk SSR, Vsesojus. Inst. Nauč. i Tehn. Informackk, Moscow, 1980, pp. 115–133. (Russian)

2. F. Bellissima, *On the modal logic corresponding to diagonalizable algebra theory*, Boll. Un. Math. Ital. (5) **15-B** (1978), 915–930.

3. C. Bernardi, *The fixed-point theorem for diagonalizable algebras*, Studia Logica **34** (1975), 239–251.

4. _____, *On the equational class of diagonalizable algebras*, Studia Logica **34** (1975), 322–331.

5. _____, *The uniqueness of the fixed-point in every diagonalizable algebra*, Studia Logica **35** (1976), 335–343.

6. S. Feferman, *Arithmetization of metamathematics in a general setting*, Fund. Math. **49** (1960), 35–92.

7. C. F. Gardiner, *A first course in group theory*, Springer-Verlag, Heidelberg, 1980.

8. G. Gargov, *A note on the provability logics of certain extensions of Heyting's arithmetic* (to appear).

9. K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme*. I, Monatsh. Math. Phys. **38** (1931), 173–198.

10. D. Guaspari, *Partially conservative extensions of arithmetic*, Trans. Amer. Math. Soc. **254** (1979), 47–68.

11. D. Guaspari and R. M. Solovay, *Rosser sentences*, Ann. Math. Logic **16** (1979), 81–99.

12. P. Hájek, *On interpretability in set theories*. I, Comment. Math. Univ. Carolin. **12** (1971), 73–79; II, ibid. **13** (1972), 445–455.

13 _____, *On interpretability in theories containing arithmetic*. II (to appear).

14. M. Hájková, *The lattice of bi-numerations of arithmetic*. I, Comment. Math. Univ. Carolin. **12** (1971), 81–104; II, ibid., 281–306.

15. P. R. Halmos, *Algebraic logic*. I. *monadic boolean algebras*, Compositio Math. **12** (1955), 217–249.

16. C. F. Kent, *"Disorder" in lattices of binumerations*, Comment. Math. Univ. Carolin. **15** (1974), 221–224.

17. M. H. Löb, *Solution of a problem of Leon Henkin*, J. Symbolic Logic **20** (1955), 115–118.

18. A Macintyre and H. Simmons, *Gödel's diagonalisation technique and related properties of theories*, Colloq. Math. **28** (1973), 165–180.

19. R. Magari, *Problemi aperti sulle algebre diagonali*, Rend. Sem. Mat. Fis. Milano **44** (1974–1975), 75–90.

20. _____, *Metodi algebrici in teoria della dimostrazione*, Boll. Un. Math. (4) **12** (1975) 252–261.

21. _____, *The diagonalizable algebras*, Boll. Un. Math. Ital. (4) **12** (supp. fasc. 3) (1975), 117–125.

22. _____, *Representation and duality theory for diagonalizable algebras*, Studia Logica **34** (1975), 305–313.

23. _____, *On the autological character of diagonalizable algebras*, Studia Logica **35** (1976), 327–333.

24. _____, *Modal diagonalizable algebras*, Boll. Un. Math. Ital. (5) **15-B** (1978), 303–320.

25. L. Manevitz and J. Stavi, Δ_2^0 *operators and alternating sentences in arithmetic*, J. Symbolic Logic **45** (1980), 144–154.

26. M. Mirolli, *On the axiomatization of finite frames of the modal system GL*, Boll. Un. Math. Ital. (5) **17-B** (1980), 1075–1085.

27. F. Montagna, *For every n, the n-freely generated algebra is not functionally free in the equational class of diagonalizable algebras*, Studia Logica **34** (1975), 315–319.

28. _____, *On the algebraization of a Feferman's predicate*, Studia Logica **37** (1978), 221–236.

29. _____, *On the diagonalizable algebra of Peano arithmetic*, Boll. Un. Math. Ital. (5) **16-B** (1979), 795–812.

30. _____, *Interpretations of the first-order theory of diagonalizable algebras in Peano arithmetic*, Studia Logica **39** (1980), 347–354.

31. _____, *The undecidability of the first-order theory of diagonalizable algebras*, Studia Logica **39** (1980), 355–359.

32. S. Orey, *Relative interpretations*, Z. Math. Logik Grundlag. Math. **7** (1961), 146–153.

33. S. Palúch, *The lattices of numerations of theories containing Peano's arithmetic*, Comment. Math. Univ. Carolin. **14** (1973), 339–359.

34. J. B. Rosser, *Extensions of some theorems of Gödel and Church*, J. Symbolic Logic **1** (1936), 87–91.

35. G. Sambin, *Un estensione del theorema di Löb*, Rend. Sem. Math. Univ. Padova **52** (1974), 193–199.

36. _____, *An effective fixed-point theorem in intuitionistic diagonalizable algebras*, Studia Logica **35** (1976), 345–361.

37. _____, *Fixed points through the finite model property*, Studia Logica **37** (1978), 287–289.

38. _____, *Topology and categorical duality in the study of semantics for model logics* (to appear).

39. K. Segerberg, *An essay in classical modal logic*, Uppsala, 1971.

40. H. Simmons, *Topological aspects of suitable theories*, Proc. Edinburgh Math. Soc. (2) **19** (1974–1975), 383–391.

41. C. Smoryński, *The incompleteness theorems*, Handbook of Mathematical Logic (J. Barwise, ed.), North-Holland, Amsterdam, 1977, pp. 821–865.

42. _____, *Beth's theorem and self-referential sentences*, Logic Colloquium '77 (A. Macintyre. L. Pacholski, and J. Paris, eds.), North-Holland, Amsterdam, 1978.

43. _____, *Calculating self-referential statements. I: explicit calculations*, Studia Logica **38** (1979), 17–36.

44. _____, *Calculating self-referential statements*, Fund. Math. **109** (1980), 189–210.

45. _____, *Calculating self-referential statements: Guaspari sentences of the first kind*, J. Symbolic Logic **46** (1981), 329–344.

46. _____, *Fifty years of self-reference in arithmetic*, Notre Dame J. Formal Logic **22** (1981), 357–374.

47. _____, *A ubiquitous fixed point calculation*, "Interpretability"; Proceedings of the First Joint Conference on the Foundations of Mathematics Organised by the Białystok Branch of the Warsaw University and Humboldt University (Berlin) (L. Szczerba and K. Prażmowski, eds.), Doktorce 1980, Filia Uniwersytetu Warszawskiego w Białymstoku (to appear).

48. _____, *Commutativity and self-reference*, Notre Dame J. Formal Logic (to appear).

49. _____, *The finite inseparability of the first-order theory of diagonalisable algebras*, Studia Logica (to appear).

50. R. M. Solovay, *Provability interpretations of modal logic*, Israel J. Math. **25** (1976), 287–304.

51. _____, *On interpretability in set theories* (to appear).

52. S. Stefani, *On the representation of hemimorphisms between boolean algebras*, Boll. Un. Math. Ital. (5) **13-A** (1976), 206–211.

53. M. H. Stone, *The theory of representations for boolean algebra*, Trans. Amer. Math. Soc. **40** (1936), 37–111.

54. V. Švejdar, *A sentence that is difficult to interpret* (to appear).

55. A Ursini, *Intuitionistic diagonalizable algebras*, Algebra Universalis **9** (1979), 229–237.

56. _____, *A modal calculus analogous to K4W, based on intuitionistic propositional logic. I*, Studia Logica **38** (1979), 297–311.

57. A. Visser, *On the provability logic of any recursively enumerable extension of Peano arithmetic, or, another look at Solovay's proof* (to appear).

429 So. Warwick, Westmont, Illinois 60559