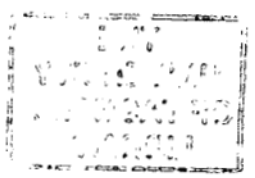
SMALL SEMIGROUP RELATED STRUCTURES WITH INFINITE PROPERTIES

Marcel Jackson (Marcel Gerard)



Submitted in fulfilment of

the requirements for the degree of

Doctor of Philosophy

University of Tasmania

June 1999

I declare that this thesis contains no material that has been accepted for the award of a degree or diploma by the University or by any other institution, and that, to the best of my knowledge and belief, it contains no material previously published or written by another person except when due reference is made in the text of the thesis.

Marcel Jackson

Marcel Jackson

Abstract

In mathematics, one frequently encounters constructions of a pathological or critical nature. In this thesis we investigate such structures in semigroup theory with a particular aim of finding small, finite, examples with certain associated infinite characteristics.

We begin our investigation with a study of the identities of finite semigroups. A semigroup (or the variety it generates) whose identities admit a finite basis is said to be *finitely based.* We find examples of pairs of finite (aperiodic) finitely based semigroups whose direct product is not finitely based (answering a question of M. Sapir) and of pairs of finite (aperiodic) semigroups that are not finitely based whose direct product is finitely based. These and other semigroups from a large class (the class of finite Rees quotients of free monoids) are also shown to generate varieties with a chain of finitely generated supervarieties which alternate between being finitely based and not finitely based. Furthermore it is shown that in a natural sense, "almost all" semigroups from this class are not finitely based.

Not finitely based semigroups that are locally finite and have the property that every locally finite variety containing them is also not finitely based are said to be *inherently not finitely based.* We construct all minimal inherently not finitely based divisors in the class of finite semigroups and establish several results concerning a fundamental example with this property; the six element Brandt semigroup with adjoined identity element, $\mathbf{B}_2^1$.

We then find the first examples of finite semigroups admitting a finite basis of identities but generating a variety with uncountably many subvarieties (indeed with a chain of subvarieties with the same ordering as the real numbers). For some well known classes, a complete description of the members with this property are obtained and related examples and results concerning joins of varieties are also found. A connection between these results and the construction of varieties with decidable word problem but undecidable uniform word problem is investigated.

Finally we investigate several embedding problems not directly concerned with semigroup varieties and show that they are undecidable. The first and second of these problems concern the fundamental relations of Green; in addition some small examples are found which exhibit unusual related properties and a problem of M. Sapir is solved. The third of the embedding problems concerns the potential embeddability of finite semigroup amalgams. The results are easily extended to the class of rings.

Acknowledgments

I am indebted to my supervisor Peter Trotter whose countless hours of discussion, suggestions and guidance helped shape this thesis. Thanks must also go to my family and to my friends for their stimulating discussions and in particular to Kelly O'Connell whose support and patience helped keep me going. Finally I would like to thank Mikhail Volkov for his useful suggestions and encouragement.

The financial support of an Australian Postgraduate Award is gratefully acknowledged.

Contents

Chapter 1

Introduction.

1.1 Historical overview

1.1.1 Critical structures

In many areas of mathematics, there exist special structures with critical properties. This critical nature can manifest itself in different ways but the importance and interest in such structures is often fundamental. In many cases the discovery of various critical structures have shaped the history of the associated theory. There are irreducible units such as the prime numbers in number theory or the finite simple groups of finite group theory, generic examples containing or mimicking the properties of other structures such as universal Turing machines and universal Diophantine polynomials, structures with pathological hereditary properties such as the graphs K_5 and $K_{3,3}$ in relation to graph planarity, or the lattices M_5 and N_5 with respect to lattice distributivity and modularity, and structures with essentially difficult properties such as aperiodic tilings, non-recursive sets of natural numbers, and flexible polyhedral surfaces.

Semigroup Theory is by no means exempt from the existence of critical constructions, in fact it is particularly rich in examples. As an example of irreducible units,

there is (aside from the celebrated congruence free (simple) groups mentioned above) the class of completely 0-simple and completely simple semigroups (irreducible with respect to the taking of Rees quotients); the structure of these is determined by the powerful results of Rees and Suschkewitz. For generic constructions there are the full transformation semigroups $\mathcal{T}_X$ and the symmetric inverse semigroups $\mathcal{I}_X$ on a set X, which respectively contain as subsemigroups all semigroups or inverse semigroups less than a certain size. Examples of structures with essentially difficult properties are the finitely presented semigroups with undecidable word problems and examples of semigroups with pathological properties include the inherently nonfinitely based semigroups.

1.1.2 Varieties, identities and Tarski's Finite Basis Problem

A very useful concept in algebra is that of a *variety*, that is, an equationally defined class of algebras. These were originally introduced and developed by G. Birkhoff [5] in 1935 who showed that a variety is equivalent to a class of algebras (of a fixed type) closed under taking direct products, subalgebras and homomorphic images. These very natural classes of algebras have been extensively investigated since their introduction and are an excellent source of critical examples. Intriguing examples include structures with some kind of *finite* character which also exhibit surprising *infinite* facets to their behaviour. Such examples can arise by examining the properties of varieties generated by finite algebras. For example, a particularly interesting aspect of a variety is the cardinality of the smallest defining set of identities. Birkhoff [5] showed that the set of identities in at most n variables satisfied by any given finite algebra can be derived from a finite subset of these identities, that is, they are finitely based. Then in 1951, R. Lyndon [49] showed that the identities of a two element algebra, of *any* type are also finitely based. At this stage there was, perhaps, some reason to suspect that the variety generated by *any* finite algebra would be finitely based. However in 1954 Lyndon [50] found a 7 element groupoid

with no finite basis for its identities. Many more examples were subsequently found including (P. Perkins, [63]) in particular the six element semigroup, $\mathbf{B}_2^1$, consisting of the matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

under matrix multiplication (alternatively $\mathbf{B}_2^1$ may be visualised as the monoid with semigroup presentation $\langle 1, a, b : aba = a, bab = b, a^2 = b^2 = 0 \rangle$). A further example presented in [63] is the Rees Quotient $\{a, b, c\}^*/I(W)$, where $I(W)$ is the ideal of the free monoid $\{a, b, c\}^*$ consisting of all words that are not subwords of a word in the set $W = \{abcba, acbab, abab, aab\}$.

While there are many known nonfinitely based finite algebras (and semigroups) there are also many well known varieties in which every finite algebra is finitely based. Some well known examples are the variety of commutative semigroups [63], the variety of idempotent semigroups ([6], [19], and [22]) and the varieties of groups [59] and rings ([44] and [47]); there are of course many others. In the 1960's, A. Tarski posed the problem of finding an algorithm to determine when a finite algebra has a finite basis for its identities. This problem, known as Tarski's Finite Basis Problem motivated much of the research into this topic and investigations gave rise (see [57] or [64]) to a new concept, that of an inherently nonfinitely based algebra. An inherently nonfinitely based algebra is a locally finite algebra whose identities have no finite basis and for which every locally finite variety containing it is also not finitely based (note that every algebra is trivially contained in a finitely based variety that is not locally finite; namely the variety defined by the empty set of identities). An inherently nonfinitely based algebra is a good example of a structure with a pathological, hereditary property.

1.1.3 Inherently nonfinitely based groupoids

In 1984, R. McKenzie [51] proved a powerful result that associates with every finite algebra of arbitrary (finite) type, a special finite groupoid with a finite basis for its identities if and only if the original algebra has a finite basis for its identities. Therefore Tarski's Finite Basis Problem can be restricted to the case of groupoids. A number of impressive results do exist for groupoids. One example is the result of Murskii [56] that "almost all" finite groupoids are finitely based; that is, the ratio of the number of finitely based groupoids of size n to the number of all groupoids of size n tends to 1 as n tends to infinity (in fact this ratio is asymptotically proportional to n^{-6}; see [57]). So in some sense there are relatively "few" nonfinitely based finite groupoids. On the other hand a result of McNulty and Shallon [53] shows that a groupoid with an identity and zero element not satisfying any nontrivial identity of the form $x \approx W(x)$, (where $W(x)$ is a groupoid term in the letter x) is either inherently nonfinitely based or a semigroup. Furthermore, results of Jezek [37] show that even among the class of groupoids satisfying nontrivial identities of the form $x \approx W(x)$, there are many inherently nonfinitely based groupoids (in fact he shows that there are idempotent, commutative, inherently nonfinitely based groupoids with only three elements). So in another sense there appear to be "few" finitely based groupoids! In 1996 Tarski's Finite Basis Problem was finally solved in the negative by R. McKenzie [52] who showed that the class of finitely based and inherently nonfinitely based finite algebras are recursively inseparable.

1.1.4 Inherently nonfinitely based semigroups

For the class of semigroups, Tarski's Finite Basis Problem remains unsolved, however there have been a number of major steps toward a positive solution. Perhaps the most notable contribution in this direction is the aesthetic description of all finite inherently nonfinitely based semigroups by M. Sapir (see [73] and [74]). We formulate

this description in the following theorem.

THEOREM 1.1.1 *[73] Let $Z_1 \equiv x_1$ and $Z_n \equiv Z_{n-1}x_nZ_{n-1}$. Then a finite semigroup is inherently nonfinitely based if and only if it satisfies no nontrivial identity of the form $Z_n \approx W$.*

Here the words Z_n are called *Zimin words* and can be considered as critical structures in their own right (see [2] and [97] for an indication of their fundamental properties with regards to the avoidability of words). An efficient algorithmic description of the same class is given in [74] as follows. Recall that the upper hypercentre of a group is the final term in the upper central series for that group.

THEOREM 1.1.2 *[74] (i) If $\mathbf{S}$ is a finite inherently nonfinitely based semigroup then for some idempotent $e \in \mathbf{S}$, $e\mathbf{S}e$ is a finite inherently nonfinitely based submonoid of $\mathbf{S}$ with identity element e.*
(ii) If $\mathbf{S}$ is a finite monoid with period d then $\mathbf{S}$ is inherently nonfinitely based if and only if for some element $a \in \mathbf{S}$ dividing an idempotent $e \in \mathbf{S}$ the elements eae and $ea^{d+1}e$ do not lie in the same coset of the maximal subgroup $\mathbf{S}_e$ of $\mathbf{S}$ containing e with respect to the upper hypercentre $\Gamma(\mathbf{S}_e)$.

It turns out that the semigroup $\mathbf{B}_2^1$ plays a surprisingly important role in the finite basis properties of finite semigroups. Firstly, the combined results of many authors (see [82] for discussion) show that every semigroup of order less than six is finitely based so $\mathbf{B}_2^1$ is as small an example as is possible of a nonfinitely based semigroup. Secondly for a very large class of finite semigroups, $\mathbf{B}_2^1$ is the minimum example of an inherently nonfinitely based semigroup. For example if the subgroups of a semigroup $\mathbf{S}$ are all nilpotent then $\mathbf{S}$ is inherently nonfinitely based if and only if $\mathbf{B}_2^1$ is contained in the variety of S (see [74]). In contrast with this however M. Sapir also constructs for any given centreless group a finite inherently nonfinitely based semigroup which does not generate a variety containing $\mathbf{B}_2^1$.

1.1.5 The lattice of subvarieties of variety

Another source of critical examples may be found by examining the lattice of subvarieties of a given variety. In the case of the lattice of all idempotent semigroup (band) varieties (see [6], [19], and [22]) a complete description has been obtained but the lattice of *all* semigroup varieties is very complicated. It is uncountable [17] and contains no anti-atoms [18]. Even the (countable) sublattice of commutative semigroup varieties contains a copy of every finite lattice [42]! There is a loose connection between the property of being nonfinitely based and generating a variety with many subvarieties: if every subvariety of a variety is finitely based (or in fact if only countably many subvarieties are nonfinitely based) then the cardinality of the lattice of subvarieties of this variety is countable. The converse however is not true: in [75] finite semigroups are constructed which each generate a variety with only a *finite* lattice of subvarieties and yet are not finitely based. Furthermore, A. N. Trahtman [92] has shown that even a finite semigroup can generate a variety with uncountably many subvarieties. The example constructed is the monoid $\mathbf{A}_2^1$ given by the matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

under matrix multiplication. The property of a semigroup generating a variety with uncountably many subvarieties is naturally inherited by every supervariety; these include varieties generated by finite semigroups that embed $\mathbf{A}_2^1$. In particular this means that there are "quite a few" finite semigroups generating varieties with uncountably many subvarieties. It is easily verified using Theorem 1.1.2 however that $\mathbf{A}_2^1$ is inherently nonfinitely based. So $\mathbf{A}_2^1$ provides no examples of *finitely based* (or even nonfinitely based, non inherently nonfinitely based) finite semigroups whose varieties have uncountably many subvarieties. Note also that the important semigroup $\mathbf{B}_2^1$ generates a *subvariety* and not a supervariety of $\mathbf{A}_2^1$ (this fact is discussed in [74]; in fact $\mathbf{B}_2^1$ is isomorphic to a subsemigroup of $\mathbf{A}_2^1 \times \mathbf{A}_2^1$).

Another related result appearing in [75] is that the set of semigroup varieties with only finitely many subvarieties is not a sublattice of the lattice of all semigroup varieties: there exist semigroups $\mathbf{S}$ and $\mathbf{T}$ each generating varieties with only finitely many subvarieties whose direct product generates a variety with infinitely many subvarieties. This is equivalent to the join of the varieties generated by $\mathbf{S}$ and $\mathbf{T}$ having infinitely many subvarieties. A further "nice" property which is not stable under direct products is that of being finitely based [75]. A simple example is by M. Volkov (see [82]). Let $\mathbf{G}$ be any nontrivial finite group and $\mathbf{A_2}$ be the semigroup $\mathbf{A_2^1}$ with the identity element removed. Both the semigroup $\mathbf{A_2}$ (see [93]) and the group $\mathbf{G}$ (see [59]) are finitely based yet their direct product is not finitely based! Conversely examples are known of nonfinitely based finite semigroups whose direct product is finitely based [75]. All these examples however depend on the presence of nontrivial subgroups and this led M. Sapir to ask whether or not there is a pair of finite finitely based aperiodic semigroups (semigroups with only trivial subgroups) whose direct product is not finitely based. In fact (see [82] for example) the class of finitely based finite semigroups is not even closed under the taking of subsemigroups and the taking of quotients, even Rees quotients (that this is true for the class of not finitely based finite semigroups follows trivially since the one element trivial semigroup is isomorphic to a quotient and a subsemigroup of every finite semigroup). So the properties of being finitely based and nonfinitely based are quite unstable. Amazingly the class of *not* inherently nonfinitely based finite semigroups *is* closed under all of these operations and therefore forms a pseudovariety (this follows from Theorems 1.1.1 or 1.1.2). A (locally finite) semigroup that is not inherently nonfinitely based has been called *weakly finitely based* in [39] and similarly it will be convenient to denote those semigroups that are both nonfinitely based and not inherently nonfinitely based as being *weakly nonfinitely based.*

Throughout the thesis we will abbreviate the phrases *finitely based, nonfinitely based, inherently nonfinitely based, weakly finitely based*, and *weakly nonfinitely based*

as FB, NFB, INFB, WFB and WNFB respectively. A related property not mentioned above is that of being *hereditarily finitely based* (or HFB). A semigroup (or variety of semigroups) has this property if it is FB and every subvariety of the variety it generates is also FB.

1.1.6 Embedding problems: varieties

Other kinds of critical structures we will investigate are those arising from embedding problems. Broadly speaking these are problems of determining when a semigroup or related structure from a class C_1 can be embedded into a semigroup from a class C_2. Embedding problems related to closure properties of varieties on their own may appear trivial: a variety is closed under the taking of subsemigroups and so clearly the class of semigroups embeddable in a semigroup from a variety V is simply itself. Associated problems however are not so trivial. For example, Theorem 1.1.2 implies that the NFB monoid $\{a,b,c\}^*/I(W)$ from [63] (see above) is WFB. This means there is a FB locally finite variety containing $\{a,b,c\}^*/I(W)$, but says nothing about what the structure of this variety is. A natural question is to ask whether $\{a,b,c\}^*/I(W)$ can be embedded in a *finite* (or even just a locally finite) finitely based semigroup of the form $X^*/I(\mathcal{V})$ for some alphabet X and set of words $\mathcal{V}$? The existence of INFB finite semigroups shows that every finite semigroup is embeddable in a finite NFB semigroup (any finite semigroup is embeddable in the direct product of itself with a finite INFB semigroup) and the construction $\mathbf{A_2} \times \mathbf{G}$ of M. Volkov (above) can be used to show that every FB finite semigroup can be embedded in a WNFB finite semigroup [82]. By definition, every WNFB finite semigroup can be embedded in a FB locally finite semigroup but it is not known if *locally finite* can be replaced by *finite* here.

1.1.7 Embedding problems: other classes

Semigroup varieties are a rich source of critical examples. However some natural embedding problems arise in non varietal situations. By right regular representations of semigroups, for example, it can be shown that every (finite) semigroup can be embedded in a (finite) regular semigroup. If we restrict ourselves to the class of inverse semigroups, the situation is more complicated but algorithms still exist which describe when a semigroup can be embedded in an inverse semigroup (see Corollary 11.15 in Volume II of [10] for a description due to B. Schein). However if we restrict ourselves further to the seemingly basic class of all Brandt semigroups (inverse semigroups with just one non-zero ideal) then the set of subsemigroups suddenly becomes very complicated. In fact this set is not recursive (Kublanovsky, see Theorem 1.3 of [25])!

One of the most studied embedding problems is that concerning the embedding of amalgams. Roughly speaking, a semigroup amalgam

$$[\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_n; \mathbf{U}]$$

is a collection of semigroups $\mathbf{S}_1$, $\mathbf{S}_2$, ... ,$\mathbf{S}_n$ each sharing a common subsemigroup $\mathbf{U}$. Clearly, a semigroup amalgam can be thought of as a special kind of partial groupoid (a set with a partially defined binary operation). In general the problem of determining when a partial groupoid can be embedded in a semigroup can be very difficult: a result of Evans (see Connection 2.2 in [39]) shows that even the problem of determining when a partial group (see Chapter 5 of this thesis for a precise definition) is embeddable in a group or finite group is undecidable. On the other hand, the corresponding embedding problem for group amalgams is very much simpler: every group amalgam is embeddable in a group (Schreier, [81]). This result was extended by T.E. Hall ([24]) when he showed that an inverse semigroup amalgam is always embeddable in an inverse semigroup. In fact a semigroup amalgam $[\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_n; \mathbf{U}]$ is always embeddable in a semigroup if $\mathbf{U}$ is an inverse semigroup

([24], [28]). There are finite inverse semigroup amalgams which are not embeddable in any finite semigroup (see page 309 of [29] for an example due to C. J. Ash) and in fact there are semigroup amalgams which are not even "partial semigroups" in the sense that there are elements x, y, and z so that $(xy)z$ and $x(yz)$ are both defined but not equal (see page 139, Volume II of [10] for an example due to Kimura).

1.2 Outline of results

In Chapter 2 we investigate the finite basis problem for the class of discrete syntactic monoids of finite languages. If $\mathbf{S}$ is a semigroup with a subset W then we define the discrete syntactic congruence ρ_W of W by $(u, v) \in \rho_W$ if and only if for any $w \in W$ and $p, q \in \mathbf{S}^1$, $puq = w \Leftrightarrow pvq = w$. Evidently ρ_W is the largest congruence on $\mathbf{S}$ for which each element of W constitutes an entire congruence class. If W is a language (that is, a subset W of a free monoid X^*) then

$$I(W) = \{w \in X^* : pwq \notin W \ \forall p, q \in X^*\}$$

is the ideal of X^* consisting of all words in X^* that are not subwords of a word in W and X^*/ρ_W is easily seen to be the Rees quotient $X^*/I(W)$. In general for a set of words W in an alphabet we denote the discrete syntactic monoid X^*/ρ_W by $S(W)$. After many of the results of this chapter were obtained, the author received a preprint entitled "On the finite basis problem for syntactic monoids of finite languages" by O. Sapir where some similar material had been independently investigated. Some of the results were then jointly refined and developed and have been combined in the forthcoming paper [34] (see also [80]).

It is shown that a very large proportion of discrete syntactic monoids are NFB. More precisely, for any given finite alphabet X and any fixed natural number k, the ratio between the number of k element sets of words of maximum length n whose discrete syntactic monoid is NFB to the number of all k element sets of words in X of maximum length n tends to 1 as n tends to infinity. Thus, in a quite natural sense,

"almost all" discrete syntactic monoids of finite languages are NFB. Furthermore every word w is a subword of a word w' at most four letters longer than w such that $S(\{w'\})$ is NFB. On the other hand it is shown that every finite set of words of length less than four letters has a FB discrete syntactic monoid. This is used to show that the smallest possible example of a NFB discrete syntactic monoid has 9 elements; a 9 element example is presented.

Another result emphasising the complicated nature of the identities of these semigroups is that for every set of words W there are finite sets of words V_i for each integer $i \geq 0$ so that $V_0 = W$, $V_i \subset V_{i+1}$ and for every $j \geq 1$, $S(V_{2j-1})$ is FB and $S(V_{2j})$ is NFB. These facts show that the class of FB (or NFB) discrete syntactic monoids of finite languages is not closed under the taking of submonoids and of Rees Quotients. A more difficult problem is that of finding finite FB (or NFB) semigroups whose direct product is NFB (or FB, respectively). Such examples have been constructed by M. Sapir [75] and M. Volkov [82] but all known examples depend on the presence of nontrivial subgroups. As noted in the introduction, an open problem of M. Sapir asks whether or not there exists a pair of FB aperiodic finite semigroups whose direct product is NFB. A solution to the dual problem of finding a pair of NFB aperiodic finite semigroups whose direct product is FB was found by O. Sapir: both $S(\{abab\})$ and $S(\{abba, aabb\})$ are NFB but their product is FB. This result is generalised and it is found that such examples are in fact quite common. We also present the first example of a pair of finite FB aperiodic semigroups whose product is NFB, answering positively the question of M. Sapir. Shortly after the discovery of this example a different example was found by O. Sapir.

In Chapter 3 we investigate the class of finite INFB semigroups. This class has been completely described by M. Sapir in [73] and [74] (see Theorems 1.1.1 and 1.1.2 above) but some interesting questions remain. It is shown in [74] that if $\mathbf{S}$ is a finite semigroup with only nilpotent subgroups then $\mathbf{S}$ is INFB if and only if $\mathbf{B_2^1} \in \mathcal{V}(\mathbf{S})$

(here and elsewhere, $\mathcal{V}(\mathbf{S})$ denotes the variety generated by the semigroup $\mathbf{S}$). We establish some similar results by showing that if C is one of the classes: finite regular semigroups; or finite semigroups whose idempotents form a subsemigroup, then a semigroup $\mathbf{S} \in C$ is INFB if and only if $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$. This is despite the fact that there are regular semigroups containing INFB subsemigroups that generate varieties not containing $\mathbf{B}_2^1$. In fact a finite regular semigroup is shown to be INFB if and only if it contains as a subsemigroup the three element semigroup consisting of the two element null semigroup with adjoined identity element. A number of corollaries are obtained which show that while every semigroup can be embedded in a regular semigroup, "very few" semigroups can be embedded in a finitely based regular semigroup. Using existing results of Rasin [71] a further corollary of the above is complete description of the FB finite orthodox monoids.

Attention is then turned to the class of minimal INFB divisors for the class of finite semigroups. Two constructions are presented for making small INFB finite semigroups whose varieties do not contain $\mathbf{B}_2^1$. Combined with the semigroups $\mathbf{B}_2^1$ and $\mathbf{A}_2^1$ and modulo certain group properties it is then shown that these form the class of minimal INFB divisors amongst finite semigroups. It is also shown that the smallest (element wise) INFB semigroup $\mathbf{S}$ for which $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$ has exactly 56 elements (all such examples are easily constructed using the given methods).

In Chapter 4 we investigate varieties with uncountable lattices of subvarieties. A. N. Trahtman [92] has shown that the finite semigroup $\mathbf{A}_2^1$ generates a variety with uncountably many subvarieties but Theorem 1.1.2 above (from [74]) shows that this variety is INFB so provides no locally finite examples of FB varieties with uncountably many subvarieties. In order to find such examples it is shown that if xyx is an isoterm for a set Σ of identities that is closed under deletion then the variety defined by Σ has uncountably many subvarieties. In fact such a variety contains a continuum of subvarieties in the sense that it contains an uncountable chain of subvarieties with the same ordering as that of the real numbers. These facts enable the con-

struction of a 7 element FB semigroup which generates a variety with uncountably many subvarieties. By combining the above with existing results, we then obtain a complete description of the varieties with uncountably many subvarieties generated by semigroups from the classes of finite orthodox monoids, INFB semigroups, discrete syntactic monoids of (not necessarily finite) languages, or monoids with index more than 2. For orthodox monoids the situation is particularly interesting since here the properties of being FB, HFB, WFB and that of generating a variety with only finitely many subvarieties are all equivalent. Likewise the properties of being NFB, INFB and generating a variety with uncountably many (semigroup) subvarieties are all equivalent for these semigroups. The theorem also enables an example to be constructed of two finitely generated varieties, one with only 3 subvarieties and the other a commutative variety with a countable infinity of subvarieties, whose join has uncountably many subvarieties. This shows that the set of varieties with only countably many subvarieties is not a sublattice of the lattice of all semigroup varieties. Several other examples of varieties with uncountably many subvarieties are investigated. In particular it is shown that if both $\mathbf{B_2}$ (the semigroup obtained from $\mathbf{B_2^1}$ by removing the identity element) and the three element monoid consisting of the two element null semigroup with adjoined identity are contained in a variety then that variety has uncountably many subvarieties. This example is used to a second 7 element WFB semigroups that generates variety with uncountably many subvarieties. The final results in Chapter 4 relate the problem of finding varieties with uncountably subvarieties to the problem of finding varieties $\mathcal{V}$ with the following unusual property: every finitely presented semigroup in $\mathcal{V}$ has a decidable word problem but there is no single algorithm which solves the word problem in any finitely presented semigroup from $\mathcal{V}$ (that is the *uniform word problem* is undecidable for $\mathcal{V}$). We show how to construct many examples of this type.

In Chapter 5 we consider some embedding problems not directly related to the study of varieties and show that these are undecidable. The first problem concerns

Green's relations $\mathcal{L}$, $\mathcal{R}$, $\mathcal{H}$, $\mathcal{D}$ and $\mathcal{J}$. These equivalence relations are some of the most useful constructions in semigroup theory, providing insight into the structure of ideals and the behaviour of subgroups with respect to other elements of a semigroup. There exists a well known and algorithmic characterisation of when a semigroup $\mathbf{S}$ with subset A can be embedded in a semigroup (or finite semigroup) $\mathbf{T}$ so that A lies within an $\mathcal{L}$- or $\mathcal{R}$-class of $\mathbf{T}$ respectively (see [21], [48] or [62] for example). Such a subset is said to be potentially $\mathcal{L}$- or potentially $\mathcal{R}$-related. The corresponding problem for the relations $\mathcal{H}$, $\mathcal{D}$ and $\mathcal{J}$ is quite different. Every semigroup is embeddable in an infinite semigroup with just one $\mathcal{D}$ and one $\mathcal{J}$ class [10] however when restricted to the class of finite semigroups the problem becomes undecidable (S. Kublanovsky, see [25]) even if the subset A consists of just two elements [45]. Likewise, we show that for the relation $\mathcal{H}$, the corresponding problem is undecidable in both the class of finite semigroups (answering problem 1 of [76]) and in the class of all semigroups, extending related results obtained by M. V. Sapir in [76]. We also show that there is no algorithm that determines when given two disjoint subsets A and B of a finite semigroup $\mathbf{S}$ whether or not $\mathbf{S}$ is embeddable in a semigroup or finite semigroup $\mathbf{T}$ so that A lies in an $\mathcal{L}$-class of $\mathbf{T}$ and B lies in an $\mathcal{R}$-class of $\mathbf{T}$. An infinite semigroup with a potentially $\mathcal{L}$- and potentially $\mathcal{R}$-related subset never lying in a $\mathcal{H}$-class of any embedding semigroup is known and in [76], the existence of a finite semigroup with this property is established. We present two eight element examples of such semigroups as well as other examples satisfying related properties.

In the final section we address embedding problems concerning finite semigroup amalgams. The most basic question to ask of a semigroup amalgam is whether or not it can be embedded in a semigroup. In general for a class C of semigroups, we will define the *strong decision problem for amalgam embeddability in* C to be the problem of determining if an amalgam of finite semigroups from C can be embedded in a semigroup from C. Similarly we define the *weak decision problem for amalgam embeddability in* C to be the problem of determining when an amalgam of finite

semigroups can be embedded in a semigroup from C. For some important classes C (such as the class of all groups [81] and the class of all inverse semigroups ([28], [24])) every finite amalgam can be embeddable in a semigroup from C and so the strong decision problem for amalgam embeddability has a very simple solution. In general however we show that the strong decision problem (and the weak decision problem) for amalgam embeddability in the class of all semigroups and the class of finite semigroups is undecidable. Furthermore the weak decision problem for amalgam embeddability in the class of inverse semigroups and the class of finite inverse semigroups is shown to be undecidable. A semigroup amalgam can be transformed into a ring amalgam by using the notion of semigroup rings. Thus a corresponding undecidability result is also obtained for the embeddability of ring amalgams into rings and finite rings.

The case of the undecidability of the decision problem for amalgam embeddability in the class of finite semigroups follows from a modification of the main result of [45]. In this paper it is shown that there is no algorithm which, when given two elements a and b of a finite semigroup $\mathbf{S}$, determines if there is a bigger finite semigroup $\mathbf{T}$ containing S in which a divides b. Using the construction of [45] it is not hard to construct an amalgam which enforces the condition a divides b in any embedding semigroup. The proof of the result of [45] however depends strongly on the rigid structure of the finite 0-simple semigroups and so this cannot be extended to the class of all semigroups (since *every* semigroup is embeddable in an infinite 0-simple semigroup in which every pair of elements divide one another). A different construction is therefore required to prove the general result. Subsequent to obtaining the results of this section, the author was informed by M. Sapir that he had earlier obtained similar results using a different method involving Minsky machines [77]. The method used by M. Sapir for the undecidability of the decision problem for amalgam embeddability in the class of finite semigroups is similar to the one we present here.

We note that some of the results in this thesis have been published, accepted for publication, or have been submitted for publication. Specifically: the results of Sections 5.2 and 5.3 appear in [30]; some of the results from Chapter 2 are to appear in [34]; and some of the results from Chapter 4 and Section 5.4 have been submitted for publication ([31], [33] and [32]).

1.3 Preliminaries: notations and definitions

In this section we define many of the basic concepts and results to be used in following chapters. In much of what follows we formulate for semigroups, concepts that also apply in a more general setting. The first reason for this restriction is because semigroups are the main concern of this thesis and the second is because in several cases slight simplifications occur under this restriction. For further general information regarding varieties and equational logic, [9] is an excellent reference. For a survey of many results specifically regarding identities of semigroups the reader is referred to [82]. There are also a number of suitable books providing information on general theory of semigroups ([10] and [29] are two of many examples). Chapter-specific notations and definitions may not appear in this section but will instead be introduced as the need arises.

A semigroup $\mathbf{S}$ consists of a set S and a binary operation $S \times S \to S$ which is associative. More formally the semigroup $\mathbf{S}$ may denoted by the pair $\langle S, \cdot \rangle$ where $\cdot$ is a symbol corresponding to a binary operation defined on the set S. In general we will relax the need for this formality and take statements such as "for every $s \in \mathbf{S}$" to mean "for every $s \in S$". An exception to this rule will be in a few definitions below and, especially, in Chapter 5 where it is beneficial to introduce greater rigour with regards to the distinction between sets and various operations defined on those sets.

1.3.1 Identities

The *free monoid* and *free semigroup* on an alphabet X will be denoted X^* and X^+ respectively. Elements of X^+ will be referred to as *words* and elements of X^* will be referred to as *possibly empty words*. The equality relation on a free monoid will be denoted "$\equiv$" and the *length* of a word w will be the number of (not necessarily distinct) letters appearing in w (denoted $|w|$). Likewise, if $W = \{w_1, \ldots, w_n\}$ is a finite set of words then the length of W is the maximum of the lengths of the words $w_1, \ldots w_n$ (denoted $|W|$). Many of the arguments to follow involve investigations into the structure of various words and for this purpose it will be convenient to introduce some notation.

DEFINITION 1.3.1 *(i) If x is a letter and w is a word, then $occ(x, w)$ is the number of occurrences of x in w,*
(ii) $c(w) = \{x : occ(x, w) > 0\}$, that is, the content of w,
(iii) a letter x is n-occurring *in a word w if $occ(x, w) = n$,*
(iv) a letter x is more than n-occurring *in a word w if for some natural number m strictly greater than n, $occ(x, w) = m$,*
(v) a word w is n-limited *if $occ(x, w) \leq n$ for all letters x.*

In the special case when a letter t is 1-occurring in a word w we will say that t is a *linear letter* in w. Several of these definitions may also be extended to finite sets of words. In particular, if $W = \{w_1, \ldots, w_n\}$ is a finite set of words then $c(W) = \cup_{i=1}^{n} c(w_i)$ and W is said to be n-limited if w_i is n-limited for every $i \leq n$.

An *identity*[1] is a formal expression $u \approx v$ where u and v are words. A semigroup $\mathbf{S}$ will be said to *satisfy* $u \approx v$ (written $\mathbf{S} \models u \approx v$) if for every assignment, θ, of elements of $\mathbf{S}$ to the letters in $c(u) \cup c(v)$, $\theta(u)$ takes the same value in $\mathbf{S}$ as $\theta(v)$ (equivalently we may say $\mathbf{S}$ satisfies $u \approx v$ where u and v are words in the alphabet

[1]The definition we give of an identity differs from the standard definition since we have restricted ourselves to the case of semigroups. More accurately, what we define is a *semigroup identity*.

X if for every homomorphism θ from X^+ into $\mathbf{S}$, $\theta(u) = \theta(v)$). A set of identities will be said to be satisfied by a semigroup if every identity in the set is satisfied by the semigroup and a set of identities will be said to be satisfied by a class K of semigroups if every identity in the set is satisfied by every semigroup in K. The set of all identities in some fixed countably infinite alphabet satisfied by a semigroup $\mathbf{S}$ (or class of semigroups K) will be denoted by $Id(\mathbf{S})$ (or $Id(K)$ respectively). The notion of satisfaction may also be extended in a natural way to include sets of identities. If Σ_1 and Σ_2 are sets of identities then $\Sigma_1 \models \Sigma_2$ if for every semigroup $\mathbf{S}$, the implication $\mathbf{S} \models \Sigma_1 \Rightarrow \mathbf{S} \models \Sigma_2$ holds.

An important kind of identity is that of the form $x^i \approx x^{i+p}$. It is easily verified that every finite semigroup $\mathbf{S}$ satisfies an identity of this form and if i and p are chosen to be minimal then i is the *index* and p is the *period* of $\mathbf{S}$. If a semigroup $\mathbf{S}$ is a group then $\mathbf{S}$ satisfies the identities $x^{1+p} \approx x$ and $x^p y \approx y x^p \approx y$ and p is also said to be the *exponent* of $\mathbf{S}$. If for some n a semigroup $\mathbf{S}$ satisfies the identity $x^n \approx x^{n+1}$ then every subgroup of $\mathbf{S}$ is a trivial group and $\mathbf{S}$ is said to be *aperiodic*. If there are no natural numbers n and m so that $\mathbf{S}$ satisfies $x^n \approx x^{n+m}$ then $\mathbf{S}$ is said to be *non-periodic*.

If Σ is a set of identities then we will say that $u \approx v$ can be *derived* from Σ (written $\Sigma \vdash u \approx v$) if there is a sequence of words $u \equiv u_1, u_2, \ldots u_{n-1}, u_n \equiv v$ in an alphabet X and homomorphisms $\theta_i : X^+ \to X^+$ so that for each $i < n$, $u_i \equiv u_i' \theta(p_i) v_i'$ and $u_{i+1} \equiv u_i' \theta(q_i) v_i'$ for some possibly empty words u_i' and v_i' and some identity $p_i \approx q_i \in \Sigma$. The homomorphisms θ_i are called *substitutions* and the number $n - 1$ is called the *length* of the derivation of $u \approx v$ from Σ. By a well known theorem (the completeness theorem for equational logic) of G. Birkhoff [5] the relations $\models$ and $\vdash$ between sets of identities are in fact equivalent.

The relation $\vdash$ enables us to formally define a *basis* of identities.

DEFINITION 1.3.2 *A finite set Σ of identities is a basis for the identities of a semigroup $\mathbf{S}$ if Σ is a minimal subset of $Id(\mathbf{S})$ from which all of $Id(\mathbf{S})$ may be*

derived.

The case when the set Σ is infinite provides more difficulties since it is possible that the identities of a semigroup $\mathbf{S}$ (with no finite basis of identities) have no irreducible subset from which $Id(\mathbf{S})$ can be derived (see [91]). However it is also known that if the identities of a semigroup $\mathbf{S}$ have a finite basis then *any* infinite subset of $Id(\mathbf{S})$ that generates $Id(\mathbf{S})$ contains a finite subset that is a basis for $Id(\mathbf{S})$ (the compactness theorem for equational logic; see Chapter II, Exercise 14.10 of [9]). Thus for our purposes it will suffice to use the following definition of an infinite basis of identities.

DEFINITION 1.3.3 *An infinite set Σ of identities is a basis for the identities of a semigroup $\mathbf{S}$ if Σ is a subset of $Id(\mathbf{S})$ from which all of $Id(\mathbf{S})$ may be derived and Σ contains no finite subset which is a basis for $Id(\mathbf{S})$.*

As noted in the introduction, if a finite basis for the identities of a semigroup exists then the semigroup is said to be *finitely based* (abbreviated to FB) and otherwise it is said to be nonfinitely based (abbreviated to NFB).

A set Σ of identities will be said to be *closed under deletion* if both $\Sigma \vdash p \approx q \Rightarrow c(p) = c(q)$ and $\Sigma \vdash p_x \approx q_x$, where $p_x \approx q_x$ is the identity obtained by deleting every occurrence of some letter x from $p \approx q$. We will say that an identity $p \approx q$ *deletes to* or *can be deleted to* $p' \approx q'$ if there is a sequence of such deletions starting at $p \approx q$ and ending at $p' \approx q'$. A word p deletes to a word p' if $p \approx p$ deletes to $p' \approx p'$. We will often be considering the semigroup identities of monoids (semigroups with identity elements). If $\mathbf{S}$ is a monoid for which there is no word w taking the value 1 under all possible assignments of elements of $\mathbf{S}$ to the letters of w (such as the word x^n does on a group of exponent n) then the set of semigroup identities satisfied by $\mathbf{S}$ is closed under deletion since assigning the element 1 to a letter in an identity is effectively the same as deleting that letter. In fact a monoid $\mathbf{S}$ for which such a word w does exist is necessarily a group since $\mathbf{S}$ must satisfy the identities $wz \approx zw \approx z$

(where $z \notin c(w)$) and from this the identity $x^{|w|}z \approx zx^{|w|} \approx z$ may easily be derived (these define the semigroup variety consisting of all groups whose exponent divides $|w|$). In general we will denote the monoid obtained from a semigroup $\mathbf{S}$ by adjoining an identity element if it does not already have one by $\mathbf{S^1}$.

It is well known that every set of identities in an alphabet X determines a fully invariant congruence on the free semigroup X^+ (that is a congruence invariant under all homomorphisms of X^+ into itself).

DEFINITION 1.3.4 *A word $p \in X^+$ is an* isoterm *relative to a set of identities Σ if $\Sigma \vdash p \approx q \Rightarrow p \equiv q$, that is, if the equivalence class of p under the fully invariant congruence corresponding to Σ is $\{p\}$. When referring to a specific semigroup $\mathbf{S}$, a word will be said to be an isoterm for $\mathbf{S}$ if it is an isoterm for $Id(\mathbf{S})$, the set of all identities satisfied by $\mathbf{S}$ over some fixed countably infinite alphabet.*

As will be seen later in this thesis, many properties of semigroup identities can determined by examining the isoterms of a semigroup.

Several of the concepts in Definition 1.3.1 are easily extended to identities.

DEFINITION 1.3.5 *(i) A letter is n-*occurring *in an identity $u \approx v$ if it is n-occurring in both u and v,*
*(ii) an identity $u \approx v$ is n-*limited *if both u and v are n-limited,*
(iii) an identity $u \approx v$ is said to be balanced *if for every letter x, $occ(x,u) = occ(x,v)$.*

1.3.2 Important classes and structural aspects of semigroups

The *variety* generated by a class of semigroups K is the closure of K under the taking of homomorphic images, subsemigroups and direct products or equivalently (by a well known theorem of G. Birkhoff [5]) the class of all semigroups satisfying $Id(K)$. This variety will be denoted by $\mathcal{V}(K)$. We may extend the notion of free semigroup to particular varieties of semigroups as follows: if $\mathcal{V}$ is a variety defined

by the set Σ of identities and θ is the corresponding fully invariant congruence on a free semigroup X^+ generated by X then the $\mathcal{V}$-free semigroup generated by X is the quotient X^+/θ. The subvarieties of a given variety $\mathcal{V}$ form a lattice under inclusion and because of the correspondence between fully invariant congruences on free semigroups and identities, the lattice of subvarieties of $\mathcal{V}$ is anti-isomorphic to the lattice of fully invariant congruences on a countably generated $\mathcal{V}$-free semigroup. As is standard in the literature, we will call a variety generated by a finite semigroup a *finitely generated variety* and a variety whose lattice of subvarieties is finite, a *small variety*. Note that for any semigroup $\mathbf{S}$ we have that $Id(\mathbf{S}) = Id(\mathcal{V}(\mathbf{S}))$ and so a semigroup has a finite basis of its identities if and only if the variety it generates can be defined by a finite set of identities[2].

While varieties are important classes of semigroups there are also many natural classes of semigroups that do not form semigroup varieties. Some important examples include: the class of semigroups in which for every element x there is an element y so that $xyx = x$ (*regular semigroups*); the class of regular semigroups in which the product of any two idempotents is again an idempotent (*orthodox semigroups*); the class of orthodox semigroups in which idempotents commute (*inverse semigroups*); the class of regular semigroups in which every element lies in a subgroup (*completely regular semigroups*); the class of finite semigroups in which all subgroups are trivial (*finite aperiodic semigroups*); the class of all finite semigroups. All but the last two of these classes contain non trivial subclasses which do form varieties. Furthermore all of these classes exhibit certain "variety-like" characteristics. For example, several of these classes are closed under the taking of subsemigroups, the taking of homomorphic images and the taking of finite direct products. Such classes are called *pseudovarieties*. Furthermore the class of inverse semigroups actually forms a variety if the unary operation $^{-1}$ is introduced (this is not a semigroup variety since

[2]The reader may wish to recall the definitions of WFB, WNFB, INFB and HFB semigroups and varieties on page 7 of the Historical Overview.

there are subsemigroups of inverse semigroups that are not inverse semigroups).

The classes just defined were mostly characterised in terms of structural properties. Some of the most important tools in investigating the structural aspects of a semigroup $\mathbf{S}$ are Green's relations defined as follows

$$\mathcal{L}^{\mathbf{S}} = \{(a,b) : \exists\ x,y \in \mathbf{S}^{\mathbf{1}} such\ that\ xa = b,\ yb = a\},$$

$$\mathcal{R}^{\mathbf{S}} = \{(a,b) : \exists\ x,y \in \mathbf{S}^{\mathbf{1}} such\ that\ ax = b,\ by = a\},$$

$$\mathcal{J}^{\mathbf{S}} = \{(a,b) : \exists\ w,x,y,z \in \mathbf{S}^{\mathbf{1}} such\ that\ wax = b,\ ybz = a\},$$

$$\mathcal{H}^{\mathbf{S}} = \mathcal{L}^{\mathbf{S}} \wedge \mathcal{R}^{\mathbf{S}},$$

$$\mathcal{D}^{\mathbf{S}} = \mathcal{L}^{\mathbf{S}} \circ \mathcal{R}^{\mathbf{S}} = \mathcal{R}^{\mathbf{S}} \circ \mathcal{L}^{\mathbf{S}}.$$

When there is no confusion as to what semigroup a particular relation is being defined on, the superscripts of these relations will be dropped. As an example of the usefulness of these relations we may reformulate several of the above definitions: regular semigroups (or inverse semigroups) are exactly the semigroups in which every $\mathcal{L}$ and every $\mathcal{R}$ class contains at least one idempotent (or exactly one idempotent, respectively). Similarly, for finite semigroups, the condition that $\mathcal{H}$ is the diagonal relational characterises the class of finite aperiodic semigroups (see [67] for a proof of this fact).

For a particular semigroup $\mathbf{S}$ and an element $a \in \mathbf{S}$ denote by L_a (respectively R_a, H_a, J_a, D_a) the equivalence class of $\mathcal{L}$ (resp. $\mathcal{R}$, $\mathcal{H}$, $\mathcal{J}$, $\mathcal{D}$) containing a. Two fundamental results associated with these relations are the following (the first is known as Green's Lemma; see [10] or [29]).

LEMMA 1.3.6 (Green). *Let a and b be two $\mathcal{R}$ equivalent elements of a semigroup $\mathbf{S}$ and let $s, t \in \mathbf{S}^1$ be such that $as = b$ and $bt = a$ (s, t exist by the definition of $\mathcal{R}$). Then the mappings given by $x \mapsto xs$ and $y \mapsto yt$ for $x \in L_a$, $y \in L_b$ are $\mathcal{R}$-class preserving, mutually inverse, injective mappings from L_a to L_b and from L_b to L_a respectively. The dual statement for $\mathcal{L}$ equivalent elements also holds.*

Let an element $a \in \mathbf{S}$ be called *regular* if there is an $x \in \mathbf{S}$ such that $axa = a$.

LEMMA 1.3.7 *(i) If a $\mathcal{D}$-class D of a semigroup $\mathbf{S}$ contains a regular element then every element of D is regular and D is called a regular $\mathcal{D}$-class of $\mathbf{S}$.*
(ii) If a $\mathcal{D}$-class D of a semigroup $\mathbf{S}$ is regular, then every $\mathcal{L}$-class and every $\mathcal{R}$-class in D contains an $\mathcal{H}$-class that is a subgroup of $\mathbf{S}$.

A *simple* semigroup[3] $\mathbf{S}$ is a semigroup containing no ideals other than itself (that is, no subsets I of $\mathbf{S}$ so that for every $s \in \mathbf{S}$, $sI \subseteq I$ and $Is \subseteq I$) and a 0-*simple* semigroup is a semigroup with a zero element 0 containing no ideals other than itself and $\{0\}$. Equivalently these two kinds of semigroups can be defined as those consisting of just one $\mathcal{J}$-class and one nonzero $\mathcal{J}$-class respectively. For a finite semigroup it can be shown that the relations $\mathcal{D}$ and $\mathcal{J}$ coincide and the simple semigroups and 0-simple semigroups admit a particularly convenient structural characterisation. Let $\mathbf{G}$ be a group and P be a $\Lambda \times I$ matrix whose entries $P_{\lambda,i}$ (with $(\lambda, i) \in \Lambda \times I$) are either 0 or elements of $\mathbf{G}$. If no row and no column of P consists entirely of zeros then we may define a semigroup operation on the set $I \times \mathbf{G} \times \Lambda \cup \{0\}$ by letting

$$(i, s, \lambda)(j, t, \nu) = \begin{cases} (i, sP_{\lambda,j}t, \nu) \; if \; P_{\lambda,j} \neq 0, \\ 0 \; otherwise \end{cases}$$

and

$$(i, s, \lambda)0 = 0 = 0(i, s, \lambda).$$

[3]While this is the usual definition of a simple semigroup it should not be confused with standard universal algebraic definition of a simple algebra which is an algebra with no nontrivial congruences.

and denote the resulting semigroup by $\mathcal{M}^0[\mathbf{G}, I, \Lambda, P]$. This semigroup is called a *Rees matrix semigroup with zero* and all finite 0-simple semigroups are isomorphic to a Rees matrix semigroup of some kind. If we insist that P contains no zero entries and remove the zero element from $\mathcal{M}^0[\mathbf{G}, I, \Lambda, P]$ we obtain an analogous description of the finite simple semigroups (denoted $\mathcal{M}[\mathbf{G}, I, \Lambda, P]$). In the infinite case there are simple and 0-simple semigroups that do not share this basic structure and in general we call those semigroups isomorphic to a Rees matrix semigroup with zero or a Rees matrix semigroup without zero as *completely 0-simple* and *completely simple* semigroups respectively. In the case where both I and Λ are finite it is often convenient to use the notation $\mathcal{M}[\mathbf{G}, n, m, P]$ (or $\mathcal{M}^0[\mathbf{G}, n, m, P]$) to denote the semigroup $\mathcal{M}[\mathbf{G}, I, \Lambda, P]$ (or $\mathcal{M}^0[\mathbf{G}, I, \Lambda, P]$ respectively) where $|I| = n$ and $|\Lambda| = m$.

One of the many reasons that completely simple and completely 0-simple semigroups are important in the study of finite semigroups is that to some extent, the structure of Rees matrix semigroup determines the structure of a $\mathcal{J}$ (or $\mathcal{D}$) class of a finite semigroup. Associated with every $\mathcal{J}$-class J_s of a semigroup $\mathbf{S}$ is an ideal $I_s = \mathbf{S}^1 s \mathbf{S}^1$ of $\mathbf{S}$. The *principal factor* of J_s is the Rees quotient I_s/I' where I' is a maximal ideal contained in I_s (if it exists) and is isomorphic to either a semigroup with zero multiplication or a Rees matrix semigroup with a zero. If I' does not exist then I_s is itself a Rees matrix semigroup.

Chapter 2

The finite basis problem for discrete syntactic monoids of finite languages.

In this chapter we investigate an interesting class of finite aperiodic semigroups (that is, semigroups with only trivial subgroups) whose identities are very simple to describe yet exhibit some complicated behavior. Recall the definition of the discrete syntactic monoid $S(W)$ of a language W (see pages 8 and 10). The identities of semigroups with this form have been of interest since P. Perkins [63] showed that $S(\{abcba, acbab, abab, aab\})$ is NFB. It is clear from the results in [73] and [74] however that for any finite set of words W, the semigroup $S(W)$ is not INFB. This means that there does exist a FB, locally finite variety containing $S(\{abcba, acbab, abab, aab\})$ and it is therefore natural to ask whether this FB, locally finite variety can be generated by a semigroup of the form $S(V)$ for some finite set of finite words V. More generally we may ask:

QUESTION 2.0.8 *(i) If W is a finite set of words, are there finite sets of words U, V such that $S(W \cup V)$ is FB and $S(W \cup U)$ is NFB?*

(ii) Conversely, do there exist finite sets of words W such that $S(V)$ is FB (or NFB) whenever $V \supseteq W$?

Another natural question (essentially Question 7.1 of [82]) is the following:

QUESTION 2.0.9 *For what finite sets of words W is the semigroup $S(W)$ FB?*

A partial solution to Question 2.0.9 has been obtained by O. Sapir in [80]:

THEOREM 2.0.10 *(O. Sapir, [80]) If w is an element of $\{a,b\}^*$ then $S(\{w\})$ is FB if and only if w is one of the following words: a^nb^m, b^na^m, a^nba^m, or b^nab^m for some n and m.*

This shows that for "most" words w in a two letter alphabet, $S(\{w\})$ is NFB! In Section 2.5 we extend this result by showing that for any fixed finite alphabet A (with $|A| > 1$) and fixed integer $k > 0$, almost all k element sets of words W in A, have a discrete syntactic monoid that is NFB. A similar (but not identical) measure concerning the number of FB semigroup operations (that is, FB associative binary operations) definable on an n-element set has the opposite solution: almost all semigroups are three nilpotent and are therefore FB [41].

This shows that the general solution to Question 2.0.9 is likely to be very complicated. Results from Sections 2.2, 2.3 and 2.4 for example show that for any finite set of finite words W we can find finite sets $V_1, V_2, \ldots$ of finite words with $W \subseteq V_1 \subset V_2 \subset \ldots$ such that $S(V_{2i})$ is FB and $S(V_{2i-1})$ is NFB for each $i > 0$. Furthermore every word w is a subword of a word w' at most four letters longer than w so that $S(\{w'\})$ is NFB. Thus we have a positive solution to Question 2.0.8 part (i) and consequently a negative solution to part (ii). It is also shown that there are finite sets of finite words U_1, U_2 and V_1, V_2 such that $S(U_1)$, $S(U_2)$ are FB, $S(V_1)$, $S(V_2)$ are NFB but $S(U_1 \cup U_2)$ is NFB and $S(V_1 \cup V_2)$ is FB.

We note that the discrete syntactic congruence is closely related to the well known *syntactic congruence* (see [67] for a precise definition) but while the syntactic congruence of a subset W of a semigroup $\mathbf{S}$ is the largest congruence on $\mathbf{S}$ that

saturates W (that is, for which W is a union of congruence classes), the discrete syntactic congruence of W is only the largest congruence on $\mathbf{S}$ that separates the elements of W in the sense that each element of W constitutes an entire congruence class. Thus, while the syntactic monoid of an infinite language can be finite (as it is for the so called *recognizable* languages; see [67] for example) the discrete syntactic monoid of an infinite language is necessarily an infinite semigroup. On the other hand if a language W consists of a single word then, as is easily verified, the discrete syntactic congruence for W coincides with the syntactic congruence for W. Furthermore, the syntactic monoid of a subset of a monoid $\mathbf{S}$ is always a homomorphic image of the discrete syntactic monoid of the subset.

2.1 Preliminary definitions

The proofs in this chapter generally involve an analysis of the structure of identities and consequently it is necessary to introduce some further terminology.

DEFINITION 2.1.1 *The expression* ${}_i x$ *means the* i^{th} *occurrence of a letter* x *in a word (see [34] or [80]).*

DEFINITION 2.1.2 *If* $c(w) = \{x_1, \ldots, x_n\}$ *and* $\{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}$ *(where* $m \leq n$*) is a subset of* $c(w)$ *then* $w(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$ *is the word obtained from* w *by assigning* 1 *to each of the letters in* $c(w) \backslash \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}$.

So in accordance with the definition given on page 19 we say that w *deletes to* $w(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$ and if $p \approx q$ is an identity with $c(p) = c(q) = \{x_1, \ldots, x_n\}$ then $p \approx q$ *deletes to* $p(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) \approx q(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$. Since $S(W)$ is always a monoid with zero element, $S(W) \models p \approx q$ implies that $c(p) = c(q)$ and also that every identity that $p \approx q$ deletes to is an identity satisfied by $S(W)$. Because of this, in the arguments to follow in this chapter we will tacitly assume that all sets of identities are closed under deletion.

In this chapter we will be considering words with large numbers of linear (1-occurring) letters. To simplify many of the required definitions it will be convenient to assume the convention that the letter t (or t_i for any subscript i) always denotes a distinct linear letter, even if it appears to occur more than once in a word. For example the word $x_1t_1x_1x_2x_2t_2t_3$ will be the same as $x_1tx_1x_2x_2tt$. We will use τ to denote the set of all linear letters in a word and subwords between successive linear letters in a word will be called *blocks*.

DEFINITION 2.1.3 *If w is a word $a_1a_2\ldots a_n$ (the a_i are not necessarily distinct letters) then $\lfloor w,t\rfloor$ is the word $a_1ta_2ta_3t\ldots a_nt$, where different occurrences of t, as usual, represent distinct linear letters. Likewise $\lfloor t,w\rfloor$ is the word $ta_1ta_2ta_3t\ldots a_n$.*

DEFINITION 2.1.4 *A pair of letters (x,y) in an identity $p\approx q$ is called stable if $p(x,y)\equiv q(x,y)$. If (x,y) is not stable in $p\approx q$ we will say it is unstable in this identity. A pair of letters is stable in a word w with respect to a semigroup $\mathbf{S}$ if $\mathbf{S}\models w\approx v$ implies (x,y) is stable in $w\approx v$.*

Note that if (x,y) is unstable in an identity $p\approx q$ is and only if (y,x) is unstable in $p\approx q$. Naturally, if every pair of letters is stable in an identity then that identity is a tautology (trivial identity). We can define a similar notion of stability for pairs of the form $({}_ix,{}_jy)$.

DEFINITION 2.1.5 *A pair $({}_ix,{}_jy)$ is stable in an identity $u\approx v$ if the order of appearance of the i^{th} occurrence of x and the j^{th} occurrence of y is the same in both u and v. If $({}_ix,{}_jy)$ is not stable in $u\approx v$ then we will say it is unstable in this identity. An unstable pair $({}_ix,{}_jy)$ is a critical pair for $u\approx v$ if it is unstable in $u\approx v$ and $({}_ix)({}_jy)$ is a subword of u.*

The identities of discrete syntactic monoids of finite languages are easy to investigate because they are described entirely in terms of isoterms: if w is a subword of a word in W then w is an isoterm for the identities of $S(W)$. There may however be

many isoterms for a monoid $S(W)$ that are not subwords of the set W. For example xx, xxy and yxx are all isoterms for $\mathbf{S} = S(\{abb, aab\})$ since they are equivalent, up to a change in names of letters to the words bb, abb and aab, all of which are words or subwords of words in the set $\{abb, aab\}$. However xyx is also an isoterm for $\mathbf{S}$ since if $\mathbf{S}$ satisfies an identity $xyx \approx w$ for some word w then because xx is an isoterm for $\mathbf{S}$ and xyx is 2-limited, $occ(x, w) = 2$ and $occ(y, w) = 1$. Since xxy and yxx are both isoterms for $\mathbf{S}$, w must be xyx.

2.2 FB discrete syntactic monoids of finite languages

In this section we find finite bases for the discrete syntactic monoid of some sets of words. The first we consider is the set, W_n, of all n-limited words in the alphabet $\{a, b\}$.

Let $\mathcal{A}_n$ denote closure under deletion of letters of the system of two identities:

$$\{x^n \approx x^{n+1}, t_1xt_2xt_3x...t_nx \approx x^nt_1t_2...t_n\}.$$

THEOREM 2.2.1 *For each $n > 0$, $S(W_n)$ is FB.*

Proof: Let an identity $u \approx v$ be called *n-simple* if the identity obtained form $u \approx v$ by deleting all more than n-occurring letters is a tautology. We show that $Id(S(W_n))$ is exactly the set of all n-simple identities. Let $S(W_n) \models p \approx q$. If a letter x is less than $(n+1)$-occurring in p then we necessarily have $occ(x, p) = occ(x, q)$ since in this case $p(x)$ is an isoterm for $S(W_n)$ (because W_n contains a copy of $p(x)$). Let (x, y) be a pair of less than $(n+1)$-occurring letters in p. Then $p(x, y)$ is an n-limited word in two letters. Since there is a copy of all such words in W_n, $p(x, y)$ must be an isoterm for $S(W_n)$. Therefore $p(x, y) \equiv q(x, y)$, and so (x, y) is a stable pair. Therefore the identity obtained from $p \approx q$ by deleting all more than n-occurring letters is a

tautology, that is $p \approx q$ is n-simple. Conversely if $p \approx q$ is an n-simple identity then since W_n is a set of n-limited words, $S(W_n)$ must satisfy $p \approx q$ because in order that p and q do not take the value 0 in $S(W_n)$ the element 1 must be assigned to every more than n-occurring letter. But then $p \approx q$ is reduced to a tautology.

Now we show that $\mathcal{A}_{n+1}$ is a basis for the identities of $S(W_n)$. Firstly $S(W_n) \models \mathcal{A}_{n+1}$ since $\mathcal{A}_{n+1}$ consists of n-simple identities. If $p \approx q$ is a nontrivial n-simple identity then we may repeatedly apply $t_1xt_2xt_3x...t_nxt_{n+1}x \approx x^{n+1}t_1t_2...t_{n+1}$ (or an identity obtained from this by deleting some linear letters) to every more than n-occurring letter in p until we have a word with an initial segment consisting entirely of more than n-occurring letters and with the remaining portion being n-limited. We may then use this identity (or an identity obtained from this by deleting some linear letters) again to rearrange the more than n-occurring letters in the initial segment into some alphabetical ordering. Applications of $x^{n+2} \approx x^{n+1}$ can then be applied to reduce the number of occurrences of these letters to $n+1$. Call the resulting word p'. We can do the same for the word q and derive $q \approx q'$. Since $p \approx q$ is n-simple, we have $p' \equiv q'$ and therefore $\mathcal{A}_{n+1} \vdash p \approx q$. □

Several simple corollaries follow.

COROLLARY 2.2.2 *Let* $\mathbf{S}$ *be a monoid satisfying* $\mathcal{A}_{n+1}$ *for some* $n > 0$. *Then the identities* $\mathcal{A}_{n+1}$ *are a finite basis for* $Id(\mathbf{S} \times S(W_n))$.

COROLLARY 2.2.3 *If every word in* W_n *is an isoterm for the identities of a monoid* $\mathbf{S}$ *and* $\mathbf{S} \models \mathcal{A}_{n+1}$ *then* $\mathbf{S}$ *satisfies the same identities as* $S(W_n)$ *and therefore is FB.*

COROLLARY 2.2.4 *Let* $\mathbf{S}$ *be a semigroup (or finite semigroup) satisfying the set of identities* $\mathcal{A}_n$ *for some* n. *Then* $\mathbf{S}$ *is a subsemigroup of a FB semigroup (or a FB, finite semigroup respectively).*

A semigroup is said to be k-nilpotent if the product of any k elements is 0 and a monoid is said to be k-nilpotent if it is a k-nilpotent semigroup with adjoined

identity element. It is clear that if $\mathbf{S}$ is a (finite) k-nilpotent monoid then $\mathbf{S}$ satisfies the conditions of Corollary 2.2.4, with $n = k$ and so is a subsemigroup of a finitely based (finite) semigroup. However the direct product of $\mathbf{S}$ with $S(W_k)$ is not a nilpotent semigroup (it has identity element $(1,1)$ but $(1,0)$ is also an idempotent). An alternative construction is as follows. Since $\mathbf{S}$ and $S(W_k)$ are nilpotent monoids, $\bar{\mathbf{S}} = \mathbf{S}\backslash\{1\}$ and $\bar{\mathbf{S}}(W_k) = S(W_k)\backslash\{1\}$ are nilpotent semigroups. Now consider the semigroup $\bar{\mathbf{T}}$ on the set

$$(\bar{\mathbf{S}}\backslash\{0\}) \cup (\bar{S}(W_k)\backslash\{0\}) \cup \{0\}$$

with multiplication within the subsets $\bar{\mathbf{S}}$ and $\bar{S}(W_k)$ unchanged and all other products equaling zero (this construction is called the 0-*direct join* of $\bar{\mathbf{S}}$ with $\bar{S}(W_k)$). Finally let $\mathbf{T}$ be the semigroup $\bar{\mathbf{T}}$ with adjoined identity element. It is clear that $\mathbf{T}$ contains both $\mathbf{S}$ and $S(W_k)$ as submonoids and that $\mathbf{T}$ is a $(2k+1)$-nilpotent monoid (since the longest word in W_k is $2k$ letters long). Finally Corollary 2.2.3 shows that $\mathbf{T}$ is FB. Thus we have shown the following

COROLLARY 2.2.5 *The pseudovariety generated by the class of finite, FB, nilpotent monoids (that is, the closure of this class under taking subsemigroups, homomorphic images and finite direct products) contains all finite nilpotent monoids and finite nilpotent semigroups.*

The next result uses the fact that the words in W_n are capable of "dominating" smaller collections of words.

COROLLARY 2.2.6 *If W is a finite set of words then there is a set of words $W' \supseteq W$ involving no more than $|c(W)|$ letters such that $S(W')$ is finitely based.*

Proof: If W is a finite set of words in one letter, then $S(W)$ is commutative and therefore already finitely based (see [63]). Assume then that $c(W)$ contains two letters a and b. Let n be the maximal number of times a letter appears in words in

W and take W' to be the union of W and W_n. Then any word in W_n is an isoterm for $S(W')$ and $S(W')$ satisfies $\mathcal{A}_n$. By Corollary 2.2.3, $S(W')$ is FB. □

We now examine the bases of identities of "small" sets of words.

PROPOSITION 2.2.7 *Let W be a set of words whose length is at most three. Then $S(W)$ is FB.*

Proof: We will essentially use a case by case analysis. First note that if u is a subword of a word v then $S(\{v\})$ is equationally equivalent to $S(\{u, v\})$. Furthermore, if w is the word v written in a different alphabet, then $S(\{u, w\})$ is equationally equivalent to $S(\{u, v\})$. Finally note that the word xyz is an isoterm for the $S(W)$ whenever the word xy is. This means that if W is a set of words of length at most three and W contains a word in three different letters, say abc, then the discrete syntactic monoid of the set W' obtained from W be replacing abc with ab is satisfies the same identities as $S(W)$. Thus we have (up to isomorphism and anti-isomorphism) only the following sets of words to consider:

$$\{a\}, \{ab\}, \{aa\}, \{ab, aa\}, \{aaa\},$$

$$\{ab, aaa\}, \{aab\}, \{aab, aaa\}, \{aba\}, \{aba, aa\},$$

$$\{aba, aaa\}, \{aba, aab\}, \{aab, baa\}, \{aba, aab, baa\},$$

$$\{aab, baa, aaa\}, \{aba, aab, aaa\}, \{aba, aab, baa, aaa\}.$$

In [70] it is shown that any variety satisfying the identity $xyx \approx xxy$ (or $xyx \approx yxx$) is FB. The discrete syntactic monoid of the first eight of the above cases satisfy this identity and so are FB. The last nine cases need special attention and are addressed in the following two lemmas.

LEMMA 2.2.8 *Let $\mathbf{S}_1$ be $S(\{aba\})$, $\mathbf{S}_2$ be $S(\{aba, aa\})$ and $\mathbf{S}_3$ be $S(\{aba, aaa\})$. Then a (finite) basis for $\mathbf{S}_i$ is the closure under deletion of letters of the set*

$$\{t_1xt_2x\dots t_{i+1}x \approx x^{i+1}t_1t_2...t_i, x^{i+1} \approx x^{i+2}, xyy \approx yyx,$$
$$xuyvxy \approx xuyvyx, xuyxvy \approx xuxyvy, xyuxvy \approx yxuxvy\}.$$

Proof: Let $p \approx q$ be satisfied by $\mathbf{S}_i$. It is obvious that xyx is an isoterm for $\mathbf{S}_i$ so the identity obtained by deleting all but linear letters from $p \approx q$ is a tautology. Furthermore if p (or q) deletes to xyx then so does q (or p). The identity $t_1xt_2x\dots t_{i+1}x \approx x^{i+1}t_1t_2...t_i$ may be used to move all variables occurring three or more times in p or q to one side of the word, say the left. All that remains other than these occur one or two times. Let x be a 2-occurring letter in p (or q) such that p (or q) cannot be deleted to xtx for some linear letter t. The last three identities (or identities obtained from these by deleting variables) in the above set can be used to move the two occurrences of x closer together. Eventually we obtain the subword xx and then the identity $xxy \approx yxx$ may be used to move this subword to the far left also. By repeating this for all such letters we obtain an identity $p \approx p_1$ where $p_1 \equiv w_1w_2$ and w_2 contains exactly all linear letters and all 2-occurring letters x for which p deletes to xtx for some linear letter t. The same process performed on the word q gives a similar identity $q \approx q_1$ where $q_1 \equiv v_1v_2$ and, since xyx is an isoterm for $\mathbf{S}_i$, $c(v_2) = c(w_2)$. All letters in w_1 (or v_1) are 2-occurring in w_1 (or v_1) and can be rearranged freely using the identities $x^{i+1} \approx x^{i+2}$, $t_1xt_2x\dots t_{i+1}x \approx x^{i+1}t_1t_2...t_i$, and $yxx \approx xxy$. Thus we can assume that $w_1 \equiv v_1$. It remains to show that we can derive $w_2 \approx v_2$.

Assume therefore that both p and q are 2-limited words and for every 2-occurring letter x we can delete p (and therefore q) to xtx for some linear letter t. So if $({}_ix,{}_jy)$ is a critical occurrence pair in $p \approx q$ then x and y must be 2-occurring letters. Therefore without loss of generality $p \approx q$ deletes to one of the following: $xytyzx \approx yxtyzx$, $xytyx \approx yxtxy$, $xytxzy \approx yxtxzy$, $xytxy \approx xytyx$, or $xtxyzy \approx xtyxzy$ where t is a linear letter and z is either linear or the empty word. In every case one of the

identities

$$\{xuyvxy \approx xuyvyx, xuyxvy \approx xuxyvy, xyuxvy \approx yxuxvy\}$$

can be applied to one of p to obtain an identity $p' \approx q$ where the number of unstable occurrence pairs $({}_ix,{}_jy)$ is smaller than that of $p \approx q$. Since there are only finitely many such pairs we eventually obtain a derivation of $p \approx q$ as required. □

LEMMA 2.2.9 *The closure under deletion of letters of the set*

$$\{t_1xt_2xt_3x \approx x^3t_1t_2t_3, x^3 \approx x^4,$$
$$xuyvxy \approx xuyvyx, xuyxvy \approx xuxyvy, xyuxvy \approx yxuxvy\}$$

is a (finite) basis for the identities of

$$S(\{aba, aab\}),\ S(\{aab, baa\}) \text{ and } S(\{aba, aab, baa\}).$$

Likewise, the closure under deletion of letters of set

$$\{t_1xt_2xt_3x \approx x^3t_1t_2t_3, x^4 \approx x^5,$$
$$xuyvxy \approx xuyvyx, xuyxvy \approx xuxyvy, xyuxvy \approx yxuxvy\}$$

is a (finite) basis for the identities of

$$S(\{aba, aab, aaa\}),\ S(\{aab, baa, aaa\}) \text{ and } S(\{aba, aab, baa, aaa\}).$$

After noting that xx, xyx, xxy and yxx are all isoterms for all of the semigroups in this lemma and that xxx is an isoterm for the last three semigroups, this lemma can be proved in an almost identical way to Lemma 2.2.8. □

The proof of Proposition 2.2.7 now follows. □

PROPOSITION 2.2.10 *If $S(W)$ has less than 10 elements then $S(W)$ is FB as long as W is not equivalent to $\{abab\}$ up to a change of letter names.*

Proof: If $S(W)$ is NFB, Proposition 2.2.7 implies that W must contain a word with at least four letters. It is easily verified that a word, w, of length at least four and involving three distinct letters has at least 8 distinct subwords and so $S(\{w\})$ has at least 10 elements. Now the only words w of length at least four and involving at most two distinct letters for which $S(\{w\})$ has less than 10 elements contain a subword equivalent up to a change in the names of letters to one of the words $aaaa$, $aaab$, $baaa$ and $abab$. The word $aaaa$ has only 4 distinct subwords. In [70] it is shown that if a semigroup satisfies $xyx \approx xxy$ or $xyx \approx yxx$ then it is FB. If $xxxx$ is an isoterm then in order that $S(W)$ not satisfy one of these identities, either xyx or both xyy and yyx must also be isoterms for $S(W)$. Thus W contains a word with a subword of the form uvu (where $uvu \not\equiv uuv$ and $uvu \not\equiv vuu$) or W contains words with subwords of the form uuv and $v'u'u'$ (where $uuv \not\equiv uvu$ and $v'u'u' \not\equiv u'v'u'$). It then easily follows that $S(W)$ has more than 9 elements; the smallest possibility being $S(\{aaaa, aba\})$ with 10 elements. However $S(\{aaaa, aba\})$ and $S(\{aaaa, aab, baa\})$ have at least 10 elements. The words $aaab$ and $baaa$ each have exactly 8 distinct subwords. Therefore a set of words W containing one of these words, say $aaab$, and such that $S(W)$ has at most 10 elements must be the set $W = \{aaab\}$. The proposition now follows since the semigroups $S(\{aaab\})$ and $S(\{baaa\})$ are FB by Theorem 2.0.10. □

Note that it follows from Theorem 2.0.10 above that $S(\{abab\})$ is NFB (see also Example 2.3.4 below).

We now turn our attention to one last example of a finitely based semigroup of the form $S(W)$. This example will become relevant in Section 2.6.1.

PROPOSITION 2.2.11 *The closure under deletion of letters of the set*

$$\Sigma = \{t_1xt_2xt_3x \approx x^3t_1t_2t_3, x^2 \approx x^3, xxt \approx txx, xt_1xyt_2y \approx xt_1yxt_2y\}$$

is a finite basis for the identities of $\mathbf{S} \equiv S(\{abcab, abcba\})$.

Proof: We first show that every word w can be transformed by Σ to a word of the form $x_1^2...x_n^2u$ where u is a 2-limited word not containing any of the letters $x_1, x_2, ..., x_n$ and such that for every 2-occurring letter x there is a linear letter t so that u deletes to xtx (this is similar to the process we used in Proposition 2.2.8). Let such a word be called a *reduced word* for $\mathbf{S}$.

Firstly, if x occurs more than 3 times in the word u then we may apply the identity $t_1xt_2xt_3x \approx x^3t_1t_2t_3$ to move all occurrences of it to the left. By applying $x^3 \approx x^2$ we can then reduce the number of occurrences of x to 2. Thus for any word w, $\Sigma \vdash w \approx w'$ where w' is 2-limited.

Now say that x is 2-occurring in a 2-limited word w and that there is no linear letter t in w for which $w(x,t) \equiv xtx$. So $w \equiv AxBxC$ for some words A, B and C where every letter in B is 2-occurring in w. If B is empty then we may apply $txx \approx xxt$ to move x to the left as required. If B is not empty then w is equivalent to a word of the form $AxD({}_2y)ExC$ where D contains only first occurrences of letters 2-occurring in w (this includes the situation where E is empty and x is y). We may then move y leftward out of B using repeated applications of one or both of $xt_1xyt_2y \approx xt_1yxt_2y$ and $xxt \approx txx$.

The length of B is reduced by this procedure and therefore by repeating these steps a word in which xx is a subword is eventually obtained. Both occurrences of the letter x can now be moved to the far left hand end of the word using the identity $xxt \approx txx$. Since this can be done for all 2-occurring letters x in w such that w does not delete to xtx for some t, we have shown (for some n) that $\Sigma \vdash w \approx x_1^2 \dots x_n^2u$ where u is a reduced word for $\mathbf{S}$. So if $w \approx v$ is an identity satisfied by S then we may use Σ to derive $w \approx x_1^2...x_n^2u_1$ and $v \approx x_1^2...x_n^2u_2$ where both u_1 and u_2 are reduced. Since u_1 and u_2 do not contain x_i for $i \leq n$, $\mathbf{S}$ must satisfy the identity $u_1 \approx u_2$.

In order to complete the proof we will show how Σ can be applied to reduce the number of unstable occurrence pairs $({}_ix,{}_jy)$ in an identity $u \approx v$ where u and

v are reduced words for $\mathbf{S}$. Let $u \approx v$ be such an identity of $\mathbf{S}$. Now assume that $u \approx v$ contains a critical pair of the form $({}_2x,{}_1\, y)$ or $({}_1x,{}_2\, y)$ then by applying the identity $xt_1xyt_2y \approx xt_1yxt_2y$ to the word u we obtain an identity $u' \approx v$ in which the number of unstable occurrence pairs is less than that of $u \approx v$. If $u \approx v$ contains a critical pair of the form $({}_1x,{}_1\, y)$ then without loss of generality we may assume that $u \equiv AxyBxCyD$ or $AxyByCxD$ for some words A, B, C, D. Since u is a reduced word for $\mathbf{S}$, B must contain a linear letter, t. But then we can assign a to x, b to y, c to t and 1 to all other letters and u takes the value $abcba$ or $abcab$, both of which are isoterms for $S(\{abcba, abcab\})$. This contradicts the assumption that $({}_1x,{}_1\, y)$ was a critical pair and therefore such critical pairs do not exist in $u \approx v$. The case for critical pairs of the form $({}_2x,{}_2\, y)$ follows by the symmetry of the set $\{abcba, abcab\}$.

Similarly we can show that there are no critical pairs of the form $({}_1x, t)$, $({}_2x, t)$, $(t,{}_1\, x)$, or $(t,{}_2\, x)$ (t is a linear letter as usual) since there is a linear letter between every 2-occurring letter in $u \approx v$ and xtx is an isoterm. Thus for every such (nontrivial) identity of $\mathbf{S}$, say $u \approx v$, we may always apply the identity $\{xt_1xyt_2y \approx xt_1yxt_2y\}$ to obtain an identity $u' \approx v$ with the property that $u' \approx v$ has fewer unstable pairs of the form $({}_ix,{}_j\, y)$. Since there can be only finitely many such pairs in the identity $u \approx v$, by repeating this process we eventually obtain an identity with no such pairs. This is necessarily a trivial identity and so a derivation of $u \approx v$ has been obtained. Therefore Σ is a finite basis for $S(\{abcba, abcab\})$. □

2.3 NFB discrete syntactic monoids of finite languages: background results

In this section we prove a number of nonfinite basis theorems for monoids. There will be very little interpretation or application of the results in this section; instead this will take place in Section 2.4 and Section 2.6. To begin we require some simple

results concerning isoterms.

LEMMA 2.3.1 *Let* **S** *be a monoid such that* xy *is an isoterm of* **S**. *Let* u *be an isoterm of* **S** *containing a linear letter* t_1. *Then*
(i) (O. Sapir [80]) erasing a prefix (suffix) of a block in u *gives a new isoterm for* **S** *and*
(ii) the word v *obtained by adding a linear letter* t_2 *immediately to the left (or right) of the occurrence of* t_1 *in* u *is also an isoterm for* **S**.

Proof: (ii) Let v be as in the statement of the lemma. If $v \approx w$ is a nontrivial identity satisfied by **S** then since u is an isoterm for **S**, any unstable pair in $v \approx w$ must include the letter t_2 and not the letter t_1 (note that if (t_1, t_2) was unstable then **S** would satisfy $t_1 t_2 \approx t_2 t_1$ which is not the case). Let (x, t_2) be such a pair. The word obtained from v by deleting t_1 is equivalent to u up to a change of letter names and therefore is an isoterm. This contradicts the fact that (x, t_2) is an unstable pair in $v \approx w$. Thus no such w exists and v is an isoterm for **S**. □

DEFINITION 2.3.2 *Let* $X = \{x_1, x_2, \dots\}$. *Then* $[Xn]$ *and* $[nX]$ *denote the words* $x_1 x_2 \dots x_n$ *and* $x_n x_{n-1} \dots x_1$ *respectively and* $[\mathcal{X}(2n)]$ *denotes the word*

$$x_2 x_4 \dots x_{2n} x_1 x_3 \dots x_{2n-1}.$$

We can now state and prove the first of our NFB results.

LEMMA 2.3.3 *Let* M *be an infinite set of natural numbers. If* $xyxy$ *is an isoterm for a monoid* **S** *and for every* $n \in M$, *the word* $L_n \equiv [X(2n)]t[\mathcal{X}(2n)]$ *is not an isoterm for* **S**, *then* **S** *is NFB.*

Proof: Given that $xyxy$ (and consequently $xytxy$) is an isoterm for the monoid **S** it follows that if (x_i, x_j) is an unstable pair in any identity $L_n \approx w$ satisfied by **S** then either i is even, j is odd and $j < i$ or j is even, i is odd and $i < j$. Furthermore in

this case the identity $L_n(x_i, x_j, t) \approx w(x_i, x_j, t)$ is equivalent up to a change of letter names to the identity $xytyx \approx yxtxy$. We now show that if $L_n \approx w$ is a nontrivial identity of $\mathbf{S}$, then $w \equiv [\mathcal{X}(2n)]t[X(2n)]$.

Let (x_i, x_j) be an unstable pair in a nontrivial identity $L_n \approx w$ satisfied by $\mathbf{S}$. It is convenient to denote the word to the left of t in L_n by B_1 and the word to the right of t in L_n by B_2. Since xyx is an isoterm for $\mathbf{S}$, (x_i, t) is stable in $L_n \approx w$ for any $i \leq 2n$ and so there are corresponding blocks B_1' and B_2' in w either side of the linear letter t that are permutations of the corresponding blocks B_1 and B_2 in L_n. Without loss of generality, we may assume that x_i precedes x_j in B_1 and x_j precedes x_i in B_1'. As noted above we have that i is odd, j is even and $i < j$ and therefore since $xytxy$ is an isoterm for $\mathbf{S}$ we can conclude that $w(x_i, x_j, t) \equiv x_j x_i t x_i x_j$. Now i is odd and so we have that $L_n(x_1, x_i, t) \equiv x_1 x_i t x_1 x_i$, an isoterm for $\mathbf{S}$ or $i = 1$. If i is not 1 it follows that x_1 precedes x_i in B_1' and in B_2' and also that x_1 precedes x_j in B_2' (because x_i does). As noted at the start of the proof, the pair (x_j, x_{2n}) is stable in $L_n \approx w$ and so x_{2n} occurs after x_j and therefore after x_1 in B_2'. That is, (x_1, x_{2n}) is an unstable pair in $L_n \approx w$. If x_1 precedes x_{2n} in B_1' (as it does in B_1), then $w(x_1, x_{2n}, t)$ is the word $x_1 x_{2n} t x_1 x_{2n}$, an isoterm for $\mathbf{S}$ and so contradicting the fact that (x_1, x_{2n}) was an unstable pair. So we must have x_1 occurring after x_{2n} in B_1'. Since for any odd number j', $(x_1, x_{j'})$ is stable in $L_n \approx w$, we must have $x_{j'}$ occurs after x_{2n} in B_1'. Likewise for any even number i', $x_{i'}$ precedes x_{2n} and therefore x_1 in B_1'. These facts ensure that B_1' is the word $[\mathcal{X}(2n)]$. It now easily follows that in B_2', x_1 precedes x_2, x_2 precedes x_3 and so on, so that B_2' is the word $[X(2n)]$. so $w \equiv [\mathcal{X}(2n)]t[X(2n)]$.

We now show by contradiction that if Σ is a basis for the identities of $\mathbf{S}$ then for every nontrivial identity $L_n \approx w$ satisfied by $\mathbf{S}$, Σ contains an identity with at least $2n$ letters. Since $\mathbf{S}$ satisfies such an identity for infinitely many n, this implies that Σ is infinite. If L_n is not an isoterm for $\mathbf{S}$ then we showed above that there is just one word w such that $\mathbf{S} \models L_n \approx w$. We will denote this word by R_n. So any derivation

of $L_n \approx R_n$ involves just one step. Therefore there is an identity $p \approx q \in \Sigma$ such that $L_n \equiv U_1\theta(p)U_2$ and $R_n \equiv U_1\theta(q)U_2$ (indeed it is clear from the form of $L_n \approx R_n$ that was established above that U_1 and U_2 can be taken to be empty). Say $p \approx q$ involve fewer than $2n$ distinct letters. The word $[X(2n)]$ involves $2n$ distinct letters and so there must be a letter x in $c(p)$ such that, for some $i \leq 2n-1$, x_ix_{i+1} is a subword of $\theta(x)$. This subword occurs just once in L_n and w so x must be linear in p and q. Similarly there is a variable y such that $\theta(y)$ contains a subword of $[\mathcal{X}(2n)]$ whose length is at least 2, and y is linear in p and q. However the subword $\theta(x)$ occurs before $\theta(y)$ in L_n and after $\theta(y)$ in R_n. Therefore $p(x,y) \approx q(x,y)$ is the identity $xy \approx yx$, contradicting the fact that $xyxy$ is an isoterm for $\mathbf{S}$. Hence $p \approx q$ must contain at least $2n$ distinct letters as required. Therefore $\mathbf{S}$ is NFB. □

EXAMPLE 2.3.4 *Consider $S(\{abab\})$. The word $xytxy$ is an isoterm for $S(\{abab\})$. On the other hand it is easily verified that*

$$S(\{abab\}) \models [X(2n)]t[\mathcal{X}(2n)] \approx [\mathcal{X}(2n)]t[X(2n)]$$

since for any unstable pair (z_1, z_2) in this identity, the left hand side deletes to $z_1z_2z_2z_1$ and the right hand side deletes to $z_2z_1z_1z_2$. Therefore by Lemma 2.3.3, $S(\{abab\})$ is NFB.

All our results that are based on this lemma can be proved using a similar lemma in [80]. We have included Lemma 2.3.3 for the sake of completeness and because it uses a quite different set of identities.

The following two lemmas will be useful in Section 2.6.

LEMMA 2.3.5 *Let A, B be elements of $\{xyt, yxt\}^*$ and ρ be a substitution defined by $\rho(xyt) \equiv \lfloor xy[Xn], t\rfloor$, $\rho(yxt) \equiv \lfloor [nX]yx, t\rfloor$. Let u_1, u_2, v_1, and v_2 be elements of $\{xy, yx\}$ such that u_1u_2 is not $xyxy$ and v_1v_2 is not $xyyx$.*
(a) If for some $m > 1$, Ax^my^mtB, and $AxytxytB$ are isoterms for a monoid $\mathbf{S}$ and

for every $n > 0$,

$$\mathbf{S} \models \rho(A)xyx_1^m x_2^m \dots x_n^m xyt\rho(B) \approx \rho(A)u_1 x_1^m x_2^m \dots x_n^m u_2 t\rho(B),$$

then **S** *is NFB.*

(b) If for some $m > 1$, $Ax^m y^m tB$ *and* $Axytyxt B$ *are isoterms for a monoid* **S** *and for every* n,

$$\mathbf{S} \models \rho(A)xyx_1^m x_2^m \dots x_n^m yxt\rho(B) \approx \rho(A)v_1 x_1^m x_2^m \dots x_n^m v_2 t\rho(B),$$

then **S** *is NFB.*

Proof: We will only prove part (a) since the proof of (b) is almost identical. Let L_n be the word

$$\rho(A)xyx_1^m x_2^m \dots x_n^m xyt\rho(B)$$

and R_n be the word

$$\rho(A)u_1 x_1^m x_2^m \dots x_n^m u_2 t\rho(B).$$

Let $L_n \approx w$ be a nontrivial identity satisfied by **S**. By both parts of Lemma 2.3.1, for any non-linear letter z, $L_n(z, \tau)$ is an isoterm (recall that τ is the set of linear letters in L_n; see page 28). Therefore w differs from L_n only by permutations within blocks. Since there is only one block of length more than one, the only differences between L_n and w are to be found in this block. We will refer to this block as the central block of L_n and w. Since $Ax^m y^m tB$ is an isoterm, Lemma 2.3.1 part (ii) implies that $L_n(x_i, x_j, \tau)$ is an isoterm. Thus it must be the case that $L_n(x_1, \dots, x_n, \tau)$ is an isoterm. Now $Ax^m y^m tB$ is an isoterm for **S** and so by Lemma 2.3.1 part (i), $Ax^m yytB$ and $Axxy^m tB$ are isoterms for **S**. So for any letter $x_i \in \{x_1, \dots, x_n\}$, the central block of $w(x, x_i, \tau)$ cannot be of the form xxx_i^m or $x_i^m xx$. In particular this is true for $i = 1$ and $i = n$. Likewise for any $x_i \in \{x_1, \dots, x_n\}$ the central block of $w(y, x_i, \tau)$ cannot be of the form yyx_i^m or $x_i^m yy$. Thus the central block of w is of the form $ux_1 x_2^m x_3^m \dots x_{n-1}^m x_n v$, where u is a permutation of xyx_1^{m-1} and v is a permutation of $x_n^{m-1}xy$.

Now we examine possible derivations of $L_n \approx R_n$ from the identities of **S**. In any derivation of $L_n \approx R_n$ we have a sequence of identities $I_1 \approx I_2$, $I_2 \approx I_3$, ..., $I_{k-1} \approx I_k$ such that $I_1 \equiv L_n$, $I_k \equiv R_n$ and for each i there is an identity $p_i \approx q_i$ and a substitution θ_i such that $I_i \equiv u\theta_i(p_i)v$ and $I_{i+1} \equiv u\theta_i(q_i)v$ for some words u, v. Let h smallest number such that $I_h(x,y) \not\equiv I_{h+1}(x,y)$ (this exists since by the choice of u_1 and u_2, $L_n(x,y) \not\equiv R_n(x,y)$). Both I_h and I_{h+1} are of the form of w as described above. Consider $p_h \approx q_h$. Clearly $\theta(p_h)$ contains an occurrence of x and an occurrence of y in the central block of I_h (since these occur in some different order in I_{h+1}). Therefore $\theta(p_h)$ contains both occurrences of at least one of x and both occurrences of y since otherwise the identity $\theta(p_h)(x,y) \approx \theta(q_h)(x,y)$ is the identity $xy \approx yx$, contradicting the fact that xy is an isoterm for **S**. Since the central block of both I_h and I_{h+1} contain $n+2$ distinct letters, if p_h contains less than n letters, there must be a letter z in $c(p_h)$ such that $\theta_i(z)$ contains $x_j x_{j+1}$ for some j. This subword occurs just once in I_h and I_{h+1} so z is linear in p_i. Similarly there are letters x' and y' such that $\theta_i(x')$ contains x and $\theta_i(y')$ contains y. Consider $p_h(x', y', z, \tau) \approx q_h(x', y', z, \tau)$. By the choice of I_h and I_{h+1}, the pair (x', y') is unstable in this identity. Now if z is a linear letter, $Axyzxytb$ and all subwords of this word are isoterms. Define a new substitution θ' by defining $\theta'(x') \equiv x$, $\theta'(y') \equiv y$, $\theta'(z) \equiv z$ and assigning the remaining linear letters in $p_h(x', y', z, \tau)$ to subwords of $Axyzxytb$ between corresponding occurrences of $\theta'(x')$, $\theta'(y')$ and $\theta'(z')$. That (x', y') is an unstable pair in $p_h(x', y', z, \tau) \approx q_h(x', y', z, \tau)$ now contradicts the fact that $Axyzxytb$ is an isoterm. Thus p_h must contain more than n letters. Since **S** satisfies

$$\rho(A)xyx_1^m x_2^m \ldots x_n^m xyt\rho(B) \approx \rho(A)u_1 x_1^m x_2^m \ldots x_n^m u_2 t\rho(B)$$

for every $n > 0$, any basis for $Id(\mathbf{S})$ must be infinite since for every $n > 0$ it contains an identity with more than n letters. □

EXAMPLE 2.3.6 *Consider* $S(\{abcab, abcba, a^kb^k\})$ *for some* $k > 2$. *Some isoterms for this semigroup are* $xytxyt$, $xytyxt$ *and* x^ky^kt. *On the other hand it is easy to verify that* $S(\{abcab, abcba, a^kb^k\})$ *satisfies* $xyx_1^k \ldots x_n^k xy \approx xyx_1^k \ldots x_n^k yx$. *Therefore by either part of Lemma 2.3.5,* $S(\{abcab, abcba, a^kb^k\})$ *is NFB.*

LEMMA 2.3.7 *Let* A, B *be elements of* $\{xyt, yxt\}^*$. *Say* $AxyxytB$ *and* $AyxxytB$ *are isoterms for a monoid* $\mathbf{S}$ *and for every* $n > 0$, *the word* $\sigma(A)xx[nX][Xn]t\sigma(B)$ *is not an isoterm for* $\mathbf{S}$, *where* σ *is a substitution defined by* $\sigma(xyt) \equiv \lfloor x[Xn], t \rfloor$ *and* $\sigma(yxt) \equiv \lfloor [nX]x, t \rfloor$. *Then* $\mathbf{S}$ *is NFB.*

Proof: The proof will be similar to that of the previous three lemmas. Fix some number n and let L_n be the word $\sigma(A)xx[nX][Xn]t\sigma(B)$. As in the proof of the previous lemma, Lemma 2.3.1 shows that for any nonlinear letter y in $c(L_n)$, $L(y, \tau)$ is an isoterm. Thus if $L_n \approx w$ is a nontrivial identity satisfied by $\mathbf{S}$ then w differs from L_n only by a permutation within blocks. The word $xx[nX][Xn]$ forms a block in L_n which we will refer to as the central block B_1. Since B_1 is the only block in L_n with length more than one, there is a block B_2 in w corresponding to the central block of L_n which is a permutation of $xx[nX][Xn]$. Since $AyxxytB$ is an isoterm, $L_n(x_i, x_j, \tau)$ is an isoterm for every $i, j \leq n$. Thus the central block is an interleaving of xx and $[nX][Xn]$. Because $AyxxytB$ is an isoterm for $\mathbf{S}$, the two occurrences of x in B_2 cannot lie between the two occurrences of any letter x_i since in that case $w(x, x_i, \tau)$ would be an isoterm yet (x, x_i) an unstable pair in $L_n \approx w$. Furthermore, for every $i \leq n$, the central block cannot delete to xx_ixx_i since then $w(x, x_i, \tau)$ is an isoterm and $w(x, x_i, \tau) \not\equiv L_n(x, x_i, \tau)$. Thus w is either the word

$$\sigma(A)[nX]Cxt\sigma(B),$$

where C is a interleaving of x and $[Xn]$, or the word

$$\sigma(A)x[nX][Xn]xt\sigma(B).$$

Now we show that if Σ is a set of identities with fewer than n distinct letters then $\Sigma \vdash L_n \approx w$ only if $\mathbf{S} \not\models \Sigma$. Thus any basis for $\mathbf{S}$ is infinite.

Let Σ be such a set of identities and let A' and B' be the words $A(x,\tau)$ and $B(x,\tau)$ respectively. Since $Ayxxytb$ is an isoterm, Lemma 2.3.1 implies that

$$A'xxtB' \ (\equiv (xt)^{occ(x,A)}xxt(xt)^{occ(x,B)})$$

and

$$A'xtB' \ (\equiv (xt)^{(occ(x,L_n)-1)})$$

are isoterms. Lemma 2.3.1 part (ii) implies that $A'xxztB'$ and $A'xztB'$ are also isoterms if z is a linear letter. Likewise with A'' and B'' taken to be $A(y,\tau)$ and $B(y,\tau)$ respectively it follows that $A''yxxytB''$ is an isoterm for $\mathbf{S}$. By assigning z the value xx in this word, similar arguments show that $A''yzytB''$ must be an isoterm as well. Note that up to a change in the names of letters, $A'xzxtB'$ is the word as $A''yzytB''$. Since $\Sigma \vdash L_n \approx w$ there is an identity $p \approx q \in \Sigma$ and a substitution θ such that $L_n \equiv u\theta(p)v$ and $u\theta(q)v$ is of one of the two forms derived above for w. Given the restricted nature of these two forms, $\theta(p)$ must contain the word $xx[nX][Xn]$. Now Σ contains only identities involving less than n letters so the substitution θ must assign some letter z in $c(p)$ a value containing as a subword the word x_ix_{i+1}. Since this subword occurs just once in L_n, z is linear in p. Furthermore there must be a letter x' such that $\theta(x') \equiv x$ and (x',z) is an unstable pair in $p \approx q$. In either case we have that the identity $p(x',z,\tau) \approx q(x',z,\tau)$ is not satisfied by $\mathbf{S}$ because (x',z) is an unstable pair in this identity and we can delete some linear letters so that after renaming the letter x' as x, the word $p(x',z,\tau)$ becomes a subword of one of the words $A'xxztB'$, $A'xztB'$ or $A'xzxtB'$. Thus $\mathbf{S}$ is NFB. □

EXAMPLE 2.3.8 *It is easily verified using Lemma 2.3.7 that $S(\{abab, abba\})$ is NFB.*

As mentioned (Theorem 2.0.10 above) a complete description of the finite basis property for the discrete syntactic monoids of single words in a two letter alphabet has been obtained by O. Sapir (see [80]). It turns out that the results so far obtained in this section are primarily applicable to collections of words in a two letter alphabet or in which at most two letters occur more than once. In order to address the Finite Basis Problem for more general words and sets of words it is necessary to obtain more generalised results.

First consider the following elementary lemma.

LEMMA 2.3.9 *Let w be an isoterm for a monoid* **S** *containing at least two distinct letters and X be a subset of $c(w)$. If we replace all maximal subwords of w not containing a member of X by linear letters, then the resulting word is also an isoterm for* **S**.

DEFINITION 2.3.10 *If w is a word containing the letters a and b then let $\tilde{w}$ be the word obtained from w by replacing all maximal subwords of w not containing the letters a or b by linear letters and replacing all subwords of the form ab by words of the form asb, where s is a linear letter. For example, the word $abcddbbcbababd$ would become $abt_1bbt_2bababt_3$ and then $as_1bt_1bbt_2bas_2bas_3bt_3$.*

LEMMA 2.3.11 *Let w be a word containing at least two letters a, b. If w is an isoterm for a monoid* **S** *then so is $\tilde{w}$.*

Proof: Of course (a, b) is a stable pair in $\tilde{w}$. Now let τ be the set of linear letters replacing maximal subwords of w not containing a or b and ν be the set of linear letters introduced when replacing ab by asb. As with t, we will exclude subscripts of the letter s, although different occurrences of this letter will always denote distinct linear letters. By Lemma 2.3.9, the pairs (a, t) and (b, t) are stable in $\tilde{w}$ with respect to **S** if t is from τ. Because w contains at least two letters it must contain a subword of the form xy and therefore xy is an isoterm for the monoid **S**. Thus if t_1 and t_2

are linear letters then the pair (t_1, t_2) is stable in $\tilde{w}$ with respect to $\mathbf{S}$. It remains to show that (a, s) and (b, s) are stable pairs in $\tilde{w}$ if s is a linear letter from ν.

The pair (a, s) is stable in $\tilde{w}$ because we can regain w by assigning a to a, 1 to b, maximal subwords of the form b^k to corresponding linear letters s from ν and the remaining subwords of w can be assigned to corresponding linear letters from τ. The pair (b, s) is stable in $\tilde{w}$ because we can assign b to b, 1 to a, maximal subwords of the form a^k to corresponding linear letters s and the remaining subwords of w to corresponding linear letters from τ. Therefore $\tilde{w}$ is an isoterm for $\mathbf{S}$. □

We now obtain a "general" theorem concerning the nonfinite basis properties of monoids (the vague notion of being "general" will become more precise in Section 2.5). The proof is a modified and generalised version of that used by O. Sapir to prove Theorem 2.0.10.

THEOREM 2.3.12 *Let*

$$w \equiv w_1 a^{\alpha_1} b^{\beta_1} w_2 a^{\alpha_2} p b^{\beta_2} w_3$$

be a word such that a and b are letters, p, w_1, w_2 and w_3 are possibly empty subwords and α_1, β_1, α_2 and β_2 are non zero and maximal. If both w and $xytyx$ are isoterms for a monoid $\mathbf{S}$ and for every $n \in \mathbb{N}$ the word

$$u_n \equiv \tilde{w}_1 a^{\alpha_1} [Xn] b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} t [nX] t b^{\beta_2} \tilde{w}_3$$

is not an isoterm for $\mathbf{S}$, then $\mathbf{S}$ is NFB.

Proof: As usual we will take the alphabet X to be the set $\{x_1, x_2, \dots\}$. Let $u_n \approx v_n$ be a nontrivial identity satisfied by $\mathbf{S}$. We will show that within $Id(\mathbf{S})$, identities involving arbitrarily large numbers of distinct letters are required to derive $u_n \approx v_n$ for every n. Thus no finite basis for $\mathbf{S}$ can exist, since such a basis would necessarily involve identities with a bounded number of letters. We may assume that $n > 6$.

Let Σ be a set of identities that contain less than $n - 6$ distinct letters and let $u_n \equiv p_1 \approx p_2 \approx \dots \approx p_m \equiv v_n$ be a derivation of $u_n \approx v_n$ from Σ (we may

assume that $p_1 \not\equiv p_2$). So there is an identity $u \approx v$ and words A and B such that $p_1 \equiv A\theta(u)B$ and $p_2 \equiv A\theta(v)B$ for some substitution θ. Replace the word u by the word t_1ut_2 and v by the word t_1vt_2 where t_1 and t_2 are new linear letters and extend θ by letting $\theta(t_1) \equiv A$ and $\theta(t_2) \equiv B$. So we have a derivation of $u_n \approx v_n$ from $\Sigma \cup \{t_1ut_2 \approx t_1vt_2\}$ involving at most $n-4$ letters such that $p_1 \equiv \theta(t_1ut_2)$ and $p_2 \equiv \theta(t_1vt_2)$. For the sake of simplicity, we will write simply u in place of t_1ut_2 and v in place of t_1vt_2.

Now let u' be the smallest subword of u such that $\theta(u)$ contains $[Xn]$ and u'' be the smallest subword of u such that $\theta(u'')$ contains $[nX]$. Let t be the first letter in u'. By the choice of u', $\theta(t)$ must contain x_1, the first letter of $[Xn]$. If $\theta(t)$ also contains the letter to the left of x_1 (in this case the letter a) then t must be linear in u_n, since ax_1 occurs just once in u_n. In this case, say where $\theta(t) \equiv z_1x_1z_2$ for some words z_1 and z_2 (with z_1 not empty), we can replace the letter t in u and v by the word t_3t_4 where $\theta(t_3) \equiv z_1$ and $\theta(t_4) \equiv x_1z_2$. Thus we can find a derivation of $u_n \approx v_n$ involving less than $n-3$ letters and such that $[Xn]$ is an initial segment of $\theta(u')$ (where u' is the smallest subword of u such that $\theta(u')$ contains $[Xn]$). Performing the same procedure for the end of $[Xn]$ and the start and end of $[nX]$, we can find a derivation of $u_n \approx v_n$ involving less than n letters and such that $u_n \equiv p_1 \equiv \theta(u)$, $p_2 \equiv \theta(v)$ and the smallest subword of u whose image under θ contains $[Xn]$ is assigned by θ the value $[Xn]$ and likewise for $[nX]$. We will continue with the convention that u' and v' are the smallest subwords of u so that $\theta(u') \equiv [Xn]$ and $\theta(v') \equiv [nX]$.

Since every letter in the set X occurs exactly twice in u_n or not at all, the letters occurring in u' and v' do not occur elsewhere in the word u. So because $u_n \equiv p_1 \approx p_2$ is a nontrivial identity, the word

$$I \equiv \tilde{w}_1 a^{\alpha_1} u' b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} t v' t b^{\beta_2} \tilde{w}_3$$

is not an isoterm since we can easily apply the identity $u \approx v$ to it. Our goal is to show that this contradicts the claim that w is an isoterm, thereby showing that

$\mathbf{S} \not\models \Sigma$.

Firstly, since xy is an isoterm for $\mathbf{S}$, any pair linear letters (t_1, t_2) is a stable pair in I with respect to $\mathbf{S}$. Secondly since $xytyx$ is an isoterm for $\mathbf{S}$, $[Xn]t[nX]$ is an isoterm for $\mathbf{S}$ and therefore by the choice of u' and v', the word $u'tv'$ is also an isoterm for $\mathbf{S}$. Because $u \approx v$ involves fewer than n distinct letters but $[Xn]$ and $[nX]$ each have n distinct letters, both u' and v' must contain letters t_1 and t_2 respectively such that for some $i \leq n-1$, both $\theta(t_1)$ and $\theta(t_2)$ contain the letters x_i and x_{i+1}. However every subword of $[Xn]t[nX]$ with length more than 1 occurs just once in u_n. Therefore the letters t_1 and t_2 must be linear in I. That is, both u' and v' contain a linear letter. Now if

$$w \equiv w_1 a^{\alpha_1} b^{\beta_1} w_2 a^{\alpha_2} p b^{\beta_2} w_3$$

is an isoterm for $\mathbf{S}$ then

$$\tilde{w} \equiv \tilde{w}_1 a^{\alpha_1} t b^{\beta_1} \tilde{w}_2 a^{\alpha_2} \tilde{p} b^{\beta_2} \tilde{w}_3$$

is an isoterm for $\mathbf{S}$ by Lemma 2.3.11. (Here for the sake of simplicity we are assuming that p contains at least one subword of the form ab or a letter other than a or b so that $\tilde{p}$ contains a linear letter. The only other case is when p is of the form $b^j a^k$ for some $j, k \geq 0$ and then we can replace $\tilde{p}$ in the above word by $tb^j a^k t \equiv tpt$ without effecting the arguments to follow.) By Lemma 2.3.1 part (i) the words

$$\tilde{w}_1 a^{\alpha_1} t b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} \tilde{p} b^{\beta_2} \tilde{w}_3$$

and

$$I_1 \equiv \tilde{w}_1 a^{\alpha_1} t b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} t b^{\beta_2} \tilde{w}_3$$

are also isoterms for $\mathbf{S}$. Therefore the pairs (a, b), (a, t), (b, t) are stable in I with respect to $\mathbf{S}$. If both u' and v' consist entirely of linear letters then I would be just the word I_1 with some extra linear letters placed next to existing linear letters in I_1 and therefore an isoterm by Lemma 2.3.1 part (ii), a contradiction. So let us

assume there is a letter z that is 2-occurring in $u'v'$ (all nonlinear letters in $u'v'$ are 2-occurring). To obtain the desired contradiction, it only remains to show that all linear letters t in I (not just ones that appear in $u'tv'$) and the letters a and b form stable pairs with every non linear letter z, from $u'v'$.

Let z be a 2-occurring letter in $u'v'$. For some linear letter t, I can be deleted to ztz, an isoterm for $\mathbf{S}$. If (z,s) is not stable in I for some linear letter s, then $\mathbf{S}$ must satisfy an identity $I \approx J$ with $I(s,z) \approx J(s,z)$ being the identity $zzs \approx szz$ (since zsz is an isoterm). But then $I(z,s,t) \approx J(z,s,t)$ is the identity $ztzs \approx sztz$ and so $I(s,t) \approx J(s,t)$ is the identity $st \approx ts$. This identity is not satisfied by $\mathbf{S}$ since xy is an isoterm for $\mathbf{S}$. Thus for any linear letter t in I, (z,t) is stable in I with respect to $\mathbf{S}$.

Now there is at least one linear letter in both u' and v' (say t_1 and t_3 respectively) and at least one linear letter t_2, say, between u' and v' in I. Since there exists a substitution θ such that $\theta(u') \equiv [Xn]$ and $\theta(v') \equiv [nX]$, we can choose t_1 and t_3 such that $u't_2v'$ deletes to a word of the form $zt_1t_2t_3z$ or $t_1zt_2zt_3$. Thus I can be deleted to either

$$\tilde{w}_1 a^{\alpha_1} z t_1 b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} t_2 t_3 z t_4 b^{\beta_2} \tilde{w}_3$$

or

$$\tilde{w}_1 a^{\alpha_1} t_1 z b^{\beta_1 - 1} \tilde{w}_2 a^{\alpha_2} t_2 z t_3 t_4 b^{\beta_2} \tilde{w}_3$$

Now (b,z) is stable in the first of these words since for any linear letter t, (b,t) and (z,t) are stable pairs in I and there is a linear letter t between every occurrence of b and an occurrence of z. Likewise, (a,z) is a stable pair in the second of these words.

The following assignment shows that (a,z) is also a stable pair in the first of the words: $a \to a$, $b \to 1$, $z \to b$, $t_1 \to b^{\beta_1 - 1}$, $t_2 \to p$, $t_3 \to 1$, $t_4 \to b^{\beta_2 - 1}$, and the remaining linear letters are assigned the corresponding unassigned (by the above) subwords of w. This gives the first word a value that is a subword of w and therefore an isoterm.

The following assignment shows that (b, z) is also a stable pair in the second of the words: $a \to 1$, $b \to b$, $z \to a$, $t_1 \to a^{\alpha_1 - 1}$, $t_2 \to a^{\alpha_2 - 1}$ (or $ba^{\alpha_2 - 1}$ if w_2, and therefore $\tilde{w}_2$, is empty), $t_3 \to p$, $t_4 \to 1$, and the remaining linear letters are assigned the corresponding unassigned (by the above) subwords of w. This gives the second word a value that is a subword of w and therefore an isoterm.

Since every pair of letters from $c(I)$ is stable in I with respect to $\mathbf{S}$ it must be an isoterm for $\mathbf{S}$. We have reached the desired contradiction and thus no finite basis can exist for the identities satisfied by $\mathbf{S}$. □

Note that the proof of Theorem 2.3.12 holds equally well if we replace the word apb in the statement of the theorem with bpa along with the requirement that $\tilde{p}$ contains a linear letter or equivalently, that p contains either the subword ab or a letter other than a and b (we then require that for every n,

$$\tilde{w}_1 a^{\alpha_1}[Xn]b^{\beta_1 - 1}\tilde{w}_2 b^{\beta_2} t[nX]ta^{\alpha_2}\tilde{w}_3$$

is not an isoterm). The proof also holds (after making the obvious adjustments) if the order of appearance of the two subwords ab and apb (or bpa) is reversed in the word w.

We now introduce a further definition in the style of Definition 2.3.10.

DEFINITION 2.3.13 *If w is a word then let $\ddot{w}$ be the word obtained from w by replacing every maximal subword not containing the letter a by a linear letter.*

For example if $w \equiv abcbabb$ then $\ddot{w} \equiv at_1at_2$.

Another "general" theorem is the following.

THEOREM 2.3.14 *(a) Let $w \equiv w_1u_1au_2w_2u_2au_1w_3$ be a word where a is a letter, u_1 and u_2 are non empty subwords and w_1, w_2 and w_3 are possibly empty subwords. If w is an isoterm for a monoid $\mathbf{S}$ and for every n the word*

$$r_n \equiv \ddot{w}_1 t_1[Xn]at_2\ddot{w}_2 t_3[nX]at_4\ddot{w}_3$$

is not an isoterm, then **S** *is NFB.*

(b) Let $w \equiv w_1u_1au_2w_2u_1au_2w_3$ *be a word where* a *is a letter,* u_1 *and* u_2 *are non empty subwords and* w_1, w_2 *and* w_3 *are possibly empty subwords. If* w *is an isoterm for a monoid* **S** *and for every* n *the word*

$$g_n \equiv \ddot{w}_1t_1[X2n]at_2\ddot{w}_2t_3a[\mathcal{X}2n]t_4\ddot{w}_3$$

is not an isoterm, then **S** *is NFB.*

Proof: (a) By Lemma 2.3.9 and Lemma 2.3.1 part (ii), $\ddot{w}_1t_1at_2\ddot{w}_2t_3at_4\ddot{w}_3$ is an isoterm for **S**. Therefore for any linear letter t, (a,t) is stable in r_n with respect to **S**. By assigning u_1 to x, a to y and $u_2w_2u_2$ to t in the word $xytyx$ we obtain the word $u_1au_2w_2u_2au_1$, an isoterm for **S**. Thus $xytyx$ and consequently $[Xn]t[nX]$ are isoterms for **S**. This combined with the fact that t_1t_2 is an isoterm shows that for any linear letter t, (x_i,t) is a stable pair in r_n with respect to **S**. Therefore if $r_n \approx r'_n$ is a nontrivial identity satisfied by **S**, then the only unstable pairs in $r_n \approx r'_n$ are of the form (x_i,a).

Now by the choice of u_1 and u_2, we also have that

$$\ddot{w}_1xat_1\ddot{w}_2t_2ax\ddot{w}_3$$

and

$$\ddot{w}_1ayt_1\ddot{w}_2t_2ya\ddot{w}_3$$

are isoterms for **S**. So if for some i, (x_i,a) is unstable in $r_n \approx r'_n$ then (x_i,a) is unstable in $r_n \approx r'_n$ for all i and

$$r'_n(a,x_i,\tau) \equiv \ddot{w}_1ax_it_1\ddot{w}_2t_2ax_i\ddot{w}_3.$$

Thus r'_n must be the word

$$\ddot{w}_1ta[Xn]t\ddot{w}_2ta[nX]t\ddot{w}_3.$$

Therefore any basis for $Id(\mathbf{S})$ must contain an identity $p \approx q$ with $r_n \equiv A\theta(p)B$ and $r'_n \equiv A\theta(q)B$ for some words A and B and a substitution θ. Since the only unstable pairs in $r_n \approx r'_n$ are of the form (x_i, a), we may assume that $\theta(p)$ contains both $[Xn]$ and $[nX]$. Now $[Xn]$ and $[nX]$ each contain n distinct letters and any subword of these with length more than one occurs just once in r_n. So if $p \approx q$ involves fewer than n letters then θ must assign a linear letter, t_1, in p to some subword of $[Xn]$ and a linear letter t_2 to some subword of $[nX]$. Thus (by possibly deleting some letters in $c(p)$) we find that $\mathbf{S}$ must satisfy the identity

$$\bar{r}_n \equiv \ddot{w}_1 t_1 t_5 a t_2 \ddot{w}_2 t_3 a t_6 t_4 \ddot{w}_3 \approx \ddot{w}_1 t_1 a t_5 t_2 \ddot{w}_2 t_3 t_6 a t_4 \ddot{w}_3 \equiv \bar{r}'_n.$$

However this is not possible because of the following assignment: $a \to a$, $t_5 \to u_1$, $t_6 \to 1$ and all other (linear) letters are assigned maximal unassigned portions of w. This assignment takes the word $\bar{r}_n$ to the word w but assigns $\bar{r}'_n$ the value $w_1 a u_1 u_2 w_2 u_2 a u_1 w_3$, therefore contradicting the claim that w was an isoterm. So the identity $p \approx q$ must contain at least n letters. Since r_n is not an isoterm for every n, any basis for $\mathbf{S}$ must contain infinitely many identities.

Proof: (b) As in part (a), the word $\ddot{w}_1 t_1 a t_2 \ddot{w}_2 t_3 a t_4 \ddot{w}_3$ is an isoterm for $\mathbf{S}$ and for any linear letter t, the pair (a, t) is stable in g_n with respect to $\mathbf{S}$. Now say that (x_i, x_j) is unstable in g_n with respect to $\mathbf{S}$. So (x_i, x_j) is unstable in $g_n(x_1, \ldots x_n)$ with respect to $\mathbf{S}$. Since $xytxy$ is an isoterm for $\mathbf{S}$ (because the word $u_1 a(u_2 w_2) u_1 a$ is a subword of w), Lemma 2.3.3 implies that $\mathbf{S} \models [X(2n)]t[\mathcal{X}(2n)] \approx [\mathcal{X}(2n)]t[X(2n)]$ and any basis for the identities of $\mathbf{S}$ contains an identity with at least n distinct letters. Now assume that (x_i, x_j) is a stable pair in a nontrivial identity $g_n \approx g'_n$ satisfied by $\mathbf{S}$. So for some i, (a, x_i) must be unstable in $g_n \approx g'_n$. By the choice of u_1 and u_2 the words

$$\ddot{w}_1 t_1 x_i a t_2 \ddot{w}_2 t_3 x_i a t_4 \ddot{w}_3$$

and

$$\ddot{w}_1 t_1 a x_i t_2 \ddot{w}_2 t_3 a x_i t_4 \ddot{w}_3$$

are isoterms. Thus $g'_n(a, x_i, \tau)$ must be the word $\ddot{w}_1 t a x_i t \ddot{w}_2 t x_i a t \ddot{w}_3$. Therefore since (x_i, x_j) are stable in $g_n \approx g'_n$ for all i, the pair (x_j, a) must be unstable for all j and

$$g_n(a, x_j, \tau) \equiv \ddot{w}_1 t_1 a x_j t_2 \ddot{w}_2 t_3 x_j a t_4 \ddot{w}_3.$$

So g'_n is the word $\ddot{w}_1 t_1 a[X2n] t_2 \ddot{w}_2 t_3 [\mathcal{X}2n] a t_4 \ddot{w}_3$.

Therefore if Σ is a basis for the identities of $\mathbf{S}$ then there is an identity $p \approx q \in \Sigma$ so that

$$g_n \equiv A\theta(p)B, \; g'_n \equiv A\theta(q)B$$

for some words A and B and a substitution θ. If the identity $p \approx q$ contained fewer than n letters then there must be letters z, z_1 and z_2 in p so that $\theta(z)$ contains a, $\theta(z_1)$ contains $x_i x_{i+1}$ and $\theta(z_2)$ contains $x_{2j} x_{2j+2}$ for some i, j. Evidently z_1 and z_2 are linear in $p \approx q$ and both (z_1, z) and (z_2, z) are unstable in $p \approx q$. However if we rename z as a, then both p and q are easily seen to be equivalent to a subword of the isoterm

$$\ddot{w}_1 t_1 a t_2 \ddot{w}_2 t_3 a t_4 \ddot{w}_3$$

with possibly some extra linear letters introduced next to existing linear letters. Thus a contradiction has been obtained and therefore no such identity $p \approx q$ can exist. Therefore the basis Σ must contain identities with arbitrarily large numbers of letters and is therefore infinite. □

2.4 NFB discrete syntactic monoids of finite languages

We now have all the information required to address the question as to what is the smallest semigroup $S(W)$ which is NFB. Combining Example 2.3.4 with Propositions 2.2.7 and 2.2.10 we immediately have the following theorem.

THEOREM 2.4.1 *(i) For any set of words $\{w_1, w_2, ..., w_n\}$ with the length of each w_i strictly less than 4, $S(\{w_1, ..., w_n\})$ is FB.*
(ii) If W is a set of words so that $S(W)$ has less than 10 elements then $S(W)$ is NFB if and only if $S(W) \cong S(\{abab\})$.

We note in comparison that the smallest NFB semigroup has 6 elements (take $\mathbf{B}_2^1$ for example).

Using Theorem 2.0.10 and other results in [34], [80] and above it is easy to extend Theorem 2.4.1.

THEOREM 2.4.2 *(i) For any set of words $W = \{w_1, w_2, ..., w_n\}$ with the length of each w_i strictly less than 5, $S(\{w_1, ..., w_n\})$ is FB if and only if W either contains words of each of the forms abab, abba and aabb or W does not contain a word of either of the forms abab and abba.*
(ii) If W is a set of words so that $S(W)$ has less than 11 elements then $S(W)$ is NFB if and only if either W contains a word of the form abab or abba.

Proof: (i) From Theorem 2.4.1 we need only consider sets of words of length 4. A word of length 4 that contains three distinct letters is equivalent up to a change in the names of letters to one of the words *abca*, *abac*, *aabc*, or *baac* or reverse. Each of these words can be replaced in W by perhaps several words of length at most 3 without changing the identities of $S(W)$. For example one can replace *baac* in W with the two words *baa* and *aac* (giving a new language W') since both *baa* and *aac* are isoterms for $S(W)$ and *baac* is an isoterm for $S(W')$. Therefore we need only consider the case when W contains a word equivalent up to a change of letter names to one of the words *abab*, *abba* and *aabb*. If W contains a word of the form *aabb* and not *abab* or *abba* then by a result in [80], $S(W)$ is FB. If it contains words of all three forms then it is easy to verify that W contains subwords equivalent up to a change in letter names to every 2-limited word in a two letter alphabet and therefore $S(W)$ satisfies the same identities as $S(W_2)$ and is FB by Corollary 2.2.3. Finally

we consider the case when W contains words equivalent to at least one of the words $abab$ and $abba$ but not all three of $abab$, $abba$ and $aabb$. Let W' be the subset of W containing only words of length at most 3. The cases to consider (up to a change of letter names) are $W_1 = \{abab\} \cup W'$, $W_2 = \{abba\} \cup W'$, $W_3 = \{abab, aabb\} \cup W'$, $W_4 = \{abba, aabb\} \cup W'$, and $W_5 = \{abab, abba\} \cup W'$. It is easily verified that the arguments used in Examples 2.3.8 and 2.3.4 can also be used to show that $S(W_1)$, $S(W_3)$, $S(W_5)$ are NFB. Finally consider $S(W_2)$ and $S(W_4)$. In [80] it is shown that if $\mathbf{S}$ is a monoid for which $xytyx$ is an isoterm and for every natural number n, $\mathbf{S}$ satisfies the identity $x[Xn]tx[nX] \approx [Xn]xt[nX]x$ then $\mathbf{S}$ is NFB. Using arguments similar to that in Example 2.3.4 it follows that both $S(W_2)$ and $S(W_4)$ satisfy the conditions of this result and therefore are NFB.

Proof: (ii) Given Propositions 2.2.7 and 2.2.10 we need only consider the case when $S(W)$ has 10 elements and W contains a word of length 4. As in the proof of Proposition 2.2.10 it is easily verified that every word of length 4 involving 3 distinct letters has at least 9 distinct subwords and so has a discrete syntactic monoid of at least 11 elements. Thus we need only consider the case when W contains a word in a two letter alphabet that is of length 4 or more. By symmetry it suffices to consider when W contains a word with a subword equivalent to one of the following words: $aaab$ $aaba$, $abba$, $aaaa$ $abab$. The first word has exactly 7 distinct subwords and therefore generates a discrete syntactic monoid with 9 elements. Any word w containing this as a subword must have at least 2 more subwords: w itself and at least one new subword of length 4 or less. In this case $S(\{w\})$ has more than 10 elements. Likewise for any set of words V, $S(\{aaab\} \cup V)$ has either more than 10 elements or V contains only one word v that is not a subword of $aaab$ and v is a single letter. In this case $S(\{aaab\} \cup V)$ satisfies the same identities as $S(\{aaab\})$ and is therefore also FB. The second and third words above have a discrete syntactic monoid with exactly 10 elements and are consistent with the theorem we are proving since $S(\{aaba\})$ is FB by Theorem 2.0.10 and, as mentioned above, $S(\{abba\})$ is

NFB.

Now $S(\{abab\})$ has 9 elements. Therefore if W contains $abab$ and $S(W)$ has fewer than 11 elements then either $W = \{abab\}$ or $W = \{abab, c\}$ for some letter c distinct from a and b. In both cases $S(W)$ satisfies the same identities as $S(\{abab\})$ and so is NFB. Finally if W contains a word with $aaaa$ as a subword then using the arguments of Proposition 2.2.10 we find that W must contain a word of the form aba. If the subword of the form aba involves two letters distinct from the subword of the form $aaaa$ then it follows that if $S(W)$ has more than 10 elements. The remaining case is when the subword of the form aba shares a letter with the subword of the form $aaaa$. In this case either $S(W)$ has more than 10 elements or is equivalent up to a change in letter names to $\{aaaa, aba\}$ or $\{aaaa, bab\}$, which are FB from a result in [80] (an obvious extension of Proposition 2.2.8 can also be applied). The theorem is proved. □

We will shortly apply Theorems 2.3.12 and 2.3.14 to some longer words but first it is convenient to introduce a new definition and some associated results.

If w is a word and a is a letter in $c(w)$ then we may write w as

$$w_1 a^{n_1} w_2 a^{n_2} w_3 \dots w_m a^{n_m} w_{m+1}$$

where for every $i \leq m+1$, n_i is a positive integer, w_1 and w_{m+1} are possibly empty words, w_2, w_3, ... , w_m are words and a is not contained in w_i. We may then define the *occurrence vector* of a in w to be the m-tuple $V_w(a) = (n_1, n_2, \dots, n_m)$. Clearly $\sum_{i=1}^{m} n_i = occ(a, w)$. If we replace the condition that a is a *single letter* occurring in w with the condition that a is a *subword* of w then we obtain a notion of an occurrence vector for arbitrary subwords of w. The notation $V_w(v)$ is no longer well defined however since a given subword of w may have several distinct occurrence vectors. For example the word $w \equiv aaaaa$ (where a is a letter) can be written as $(aa)^2a$ or $(aa)a(aa)$ or $a(aa)^2$ and so there are two distinct occurrence vectors for aa in w: they are (2) and (1, 1). Our primary concern will be with occurrence vectors of letters in words and for our purposes it will suffice to assume that when v is a

subword of w then $V_w(v)$ to be any one particular occurrence vector of v in w.

DEFINITION 2.4.3 *An occurrence vector*

$$v_1 = (n_1, n_2, \ldots, n_p)$$

contains *an occurrence vector*

$$v_2 = (m_1, m_2, \ldots, m_q)$$

if there is a substitution $\theta : X^* \to X^*$ with $\theta(a) \equiv a$ (for some fixed letter a) such that the word

$$a^{n_1} t_1 a^{n_2} t_2 \ldots t_{p-1} a^{n_p}$$

(where the t_i are letters) contains as a subword the word

$$\theta(a^{m_1} t_1 a^{m_2} t_2 \ldots t_{q-1} a^{m_q}).$$

In this case we will write $v_1 \geq v_2$.

For example, take v_1 and v_2 as in the definition and let $h_1, h_2, \ldots, h_q$ be a subsequence of $n_1, n_2, \ldots, n_p$ such that $m_i \leq h_i$. Consider the word

$$w \equiv a^{n_1} t_1 a^{n_2} t_2 \ldots t_{p-1} a^{n_p}.$$

Since $h_1, h_2, \ldots, h_q$ is a subsequence of $n_1, n_2, \ldots, n_p$, the word w must be of the form $w_0 a^{h_1} w_1 a^{h_2} w_2 \ldots a^{h_q} w_q$ for some words w_1, w_2, ..., w_{q-1} and some possibly empty words w_0 and w_q. Now let θ be the substitution defined by $\theta(a) \equiv a$ and $\theta(t_i) \equiv a^{h_i - m_i} w_i$. Evidently

$$\theta(a^{m_1} t_1 a^{m_2} t_2 \ldots t_{q-1} a^{m_q}) \equiv a^{h_1} w_1 a^{h_2} w_2 \ldots a^{h_q} w_q,$$

a subword of w and so by Definition 2.4.3, the occurrence vector v_1 contains the occurrence vector v_2. Also if θ is a substitution that assigns 1 to all linear letters of the form t_i in the word $w_1 \equiv a^{n_1} t_1 a^{n_2} t_2 \ldots t_{p-1} a^{n_p}$ and assigns a to itself then

$\theta(w_1) \equiv a^n$ where $n = \sum_{i=1}^{p} n_i$. Therefore the singleton occurrence vector $(n+i)$ contains the vector v_1 for any non-negative integer i. An occurrence vector of a subword u in a word w is said to be *maximal* in w if for every subword v of w, $V_w(v) \geq V_w(u) \Rightarrow u \equiv v$. Likewise if W is a set of words containing w then $V_w(u)$ is maximal in W if for every subword v of a word $w' \in W$, $V_{w'}(v) \geq V_w(u) \Rightarrow (u \equiv v$ and $w' \equiv w)$. Possibly the simplest way in which an occurrence vector $V_w(a)$ of a letter a in a word w can be maximal in a set of words W is if a occurs more times in w than any other letter and the remaining words in W are $(occ(a,w)-1)$-limited (recall Definition 1.3.1). Another simple situation is if there is a power of a in w that is higher than the power of any other subword of a word in W. On the other hand, there need not be a maximal occurrence vector amongst the set of all occurrence vectors of a word (for example in the word *aabbcc*, we have $V_w(a) = V_w(b) = V_w(c) = (2)$ and all other occurrence vectors are the singleton (1)). The importance of maximal occurrence vectors lies in the following simple lemma.

LEMMA 2.4.4 *Let w_1 and w_2 be words with u a subword of w_1 and v a subword of w_2. Let θ be a substitution. If (for some occurrence vectors $V_{w_1}(u)$ and $V_{w_2}(v)$ of u in w_1 and of v in w_2 respectively) $V_{w_1}(u) = V_{w_2}(v)$ and $V_{w_2}(v)$ is a maximal occurrence vector in a set W of words containing w_2 then $\theta(w_1)$ is a subword of a word in W only if $\theta(u) \equiv 1$ or both $\theta(u) \equiv v$ and $\theta(w_1)$ is a subword of w_2.*

Proof: This is because if $\theta(u) \not\equiv 1$ then occurrence vector of $\theta(u)$ in $\theta(w_1)$ contains the occurrence vector $V_{w_1}(u)$ which equals $V_{w_2}(v)$. Since $V_{w_2}(v)$ is maximal in W then $\theta(w_1)$ cannot be a subword of any word in W except for the word w_2 and in this case $\theta(u) \equiv v$. □

THEOREM 2.4.5 *Let W be a set of words and $w \in W$ be a word containing the letters a and b such that $V_w(a)$ is maximal in W (the set W may of course be simply $\{w\}$ itself). Let β_1 and β_2 be any positive numbers and p be any (possibly empty)*

word not containing a or b. If w satisfies one of the following conditions (or their reverse) then $S(W)$ is NFB (in each case we will assume that the given subwords of w are not contained within each other though they may overlap):

(i) w has a subword $ab^{\beta_1}a$ and a subword $apbb^{\beta_2}a$;

(ii) w has a subword $abb^{\beta_1}a$ and a subword $apb^{\beta_2}a$;

(iii) w has subwords of the form aba, apba and ba;

(iv) w contains aba and ends with apba. For example, w ends with ababa;

(v) w has a subword of the form $abbb^{\beta_1}a$ and of the form apb. For example abbbab is a subword of w;

(vi) w has a subword aba and a subword apbaa and $V_w(a)$ is the only occurrence vector of a letter in a word in W that contains the occurrence vector $V_{w'}(a)$, where w' is obtained by replacing the particular occurrence of apbaa by apaba.

Proof: In every case we will construct a set of identities $\{u_n \approx v_n\}$ based on the form of w and apply Theorem 2.3.12. Both the sides of the identities constructed will contain the letter a and in all except the last case the occurrence vectors of a in these words will be identical to that of w. Since $V_w(a)$ is maximal in W, by Lemma 2.4.4, if θ is a substitution then $\theta(u_n)$ or $\theta(v_n)$ is a subword of a word in W only if $\theta(a) \equiv 1$ or $\theta(a) \equiv a$. Furthermore, if $\theta(a) \equiv a$ then $\theta(u_n)$ (or $\theta(v_n)$) is a subword of the word w. The identities $u_n \approx v_n$ will also be constructed so that if $\theta(a) \equiv 1$ then $\theta(u_n) \equiv \theta(v_n)$. Therefore in the arguments to follow in this proof it will be sufficient to consider the case when $W = \{w\}$ and $\theta(a) \equiv a$.

First note that in every case in the theorem, w contains a subword of the form ab and another of the form ba (not intersecting). This is all that is required to establish that $xytyx$ is an isoterm for $S(\{w\})$.

We now consider each case of the theorem separately. Each of the cases involves a word w with some given subwords but the arguments we will use involving w will not depend on the order of appearance of the given subwords in w. Therefore for each case of the theorem we will only consider a particular choice for the order of

appearance of the given subwords in w.

(i) Let $w \equiv w_1({}_ia)b^{\beta_1}aw_2({}_ja)pb^{\beta_2}baw_3$, where ${}_ia$ and ${}_ja$, as usual, denote the i^{th} and j^{th} occurrences of a in w respectively.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1-1}a\tilde{w}_2({}_ja)t[nX]tb^{\beta_2}ba\tilde{w}_3$$

and

$$v_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1-1}a\tilde{w}_2({}_ja)t[nX]tb^{\beta_2}ab\tilde{w}_3.$$

Let θ be a substitution such that $\theta(u_n)$ (or $\theta(v_n)$) is a subword of w. Between ${}_ja$ and ${}_{(j+1)}a$ in w there is the word $b^{\beta_2}b$. So $\theta(t[nX]tb^{\beta_2}b)$ (or $\theta(t[nX]tb^{\beta_2})$) must be the word bb^{β_2}. Now if θ assigns b the value 1, then $\theta(u_n) \equiv \theta(v_n)$ because (a,b) is the only unstable pair in $u_n \approx v_n$. The remaining case is when $\theta(b)$ is the letter b and we will show that this never occurs ($\theta(b)$ cannot be a higher power of b since otherwise we would have more than $occ(b,w)$ occurrences of b).

If we are considering u_n then $\theta(b) \equiv b$ implies $\theta(t[Xn]t) \equiv 1$ and then the subword of u_n between ${}_ia$ and ${}_{(i+1)}a$ is simply b^{β_1-1}. Between ${}_ia$ and ${}_{(i+1)}a$ in w however, there is the word b^{β_1} and this contradicts the assumption that $\theta(u_n)$ was a subword of w since $\theta(b^{\beta_1-1})$ cannot be b^{β_1} if $\theta(b) \equiv b$. If we are considering v_n then this implies $\theta(t[nX]t) \equiv b$. If $\theta([nX]) \equiv 1$ then the previous argument applies. If $\theta(x_k) \equiv b$ for some k, then $occ(b, \theta(v_n)) > occ(b,w)$ since x_k is 2-occurring in v_n.

(ii) Let $w \equiv w_1({}_ia)bb^{\beta_1}aw_2({}_ja)pb^{\beta_2}aw_3$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1}a\tilde{w}_2({}_ja)t[nX]tb^{\beta_2}a\tilde{w}_3$$

and

$$v_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1-1}ab\tilde{w}_2({}_ja)t[nX]tb^{\beta_2}a\tilde{w}_3.$$

Let θ be a substitution such that $\theta(u_n)$ is a subword of w. Between ${}_ia$ and ${}_{(i+1)}a$ in w we have the word b^{β_1+1}. So $\theta([Xn]b^{\beta_1}) \equiv b^{\beta_1+1}$. Now $\theta(b)$ cannot be b^k for any k

greater than 1 since then $occ(b, \theta(u_n)) > occ(b, w)$. If $\theta(b) \equiv b$, then we must have $\theta(x_k) \equiv b$ for some k. But then $occ(b, \theta(u_n)) > occ(b, w)$ since x_k is 2-occurring in u_n, again contradicting the choice of θ. Thus $\theta(b) \equiv 1$, and therefore $\theta(u_n) \equiv \theta(v_n)$.

Now let $\theta(v_n)$ be a subword of w. As in the last case just considered, $\theta([Xn]b^{\beta_1-1})$ must be the word b^{β_1}. If b^{β_1-1} is empty then for some k, $\theta(x_k)$ must be the word b (since $\beta_1 = 1$). But then $\theta(t[nX]tb^{\beta_2}) \equiv pb^{\beta_2}$ since this is the word between ${}_ja$ and ${}_{(j+1)}a$ in w. Because p does not contain b and $\theta(x_k) \equiv b$, $\theta(tb^{\beta_2})$ must be b^{β_2-1}. Thus $\theta(b) \equiv 1$ and $\theta(v_n) \equiv \theta(u_n)$.

(iii) Let $w \equiv w_1({}_ia)baw_2({}_ja)pbaw_3ba\tilde{w}_4$

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2({}_ja)t[nX]tba\tilde{w}_3tba\tilde{w}_4$$

and

$$v_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2at[nX]tba\tilde{w}_3tabw_4.$$

Between ${}_ia$ and ${}_{(i+1)}a$ in w there is a single letter b. Thus $\theta([Xn]) \equiv b$. So there is an x_k such that $\theta(x_j) \equiv b$. Between ${}_ja$ and ${}_{(j+1)}a$ in w is the word pb. Therefore $\theta(t[nX]tb) \equiv pb$. Since $\theta(x_k) \equiv b$, we must have that $\theta(b) \equiv 1$ and $\theta(u_n) \equiv \theta(v_n)$.

(iv) Let $w \equiv w_1({}_ia)baw_2({}_ja)pba$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2({}_ja)t[nX]tba$$

and

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2({}_ja)t[nX]tab.$$

Let θ be such that $\theta(u_n)$ or $\theta(v_n)$ is a subword of w. As in previous cases we may deduce from the subword $({}_ia)[Xn]a$ of both u_n and v_n that $\theta(x_k) \equiv b$ for some k. So $\theta([Xn])$ is the letter b. But then from the subword $({}_ja)t[nX]tba$ in u_n and $({}_ja)t[nX]tab$ in v_n we may deduce that $\theta(b) \equiv 1$ and therefore $\theta(u_n) \equiv \theta(v_n)$.

(v) Let $w \equiv w_1({}_ia)bbb^{\beta_1}aw_2({}_ja)pbw_3$.

Claim: $S(\{w\}) \models u_n \approx v_n$, where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1}ba\tilde{w}_2at[nX]tb\tilde{w}_3$$

and

$$v_n \equiv \tilde{w}_1({}_ia)[Xn]b^{\beta_1}ab\tilde{w}_2at[nX]tb\tilde{w}_3.$$

Between ${}_ia$ and ${}_{(i+1)}a$ in w we have the word bbb^{β_1}. Between $({}_ia)$ and $({}_{(i+1)}a)$ in u_n and v_n we have $[Xn]b^{\beta_1}b$ and $[Xn]b^{\beta_1}$ respectively. Therefore if θ is a substitution such that $\theta(u_n)$ or $\theta(v_n)$ is a subword of w, then $\theta([Xn]b^{\beta_1}b)$ or $\theta([Xn]b^{\beta_1})$ respectively must be the word bbb^{β_1}. In both cases if we do not have $\theta(b) \equiv 1$, then $occ(b, \theta(u_n))$ (or $occ(b, \theta(v_n))$) is greater than $occ(b, w)$ since for all $x_k \in c([Xn])$, x_k is 2-occurring in u_n and v_n. Thus $\theta(b) \equiv 1$ and $\theta(u_n) \equiv \theta(v_n)$.

(vi) For example, $w \equiv w_1({}_ia)baw_2({}_ja)pbaaw_3$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2({}_ja)t[nX]tbaa\tilde{w}_3$$

and

$$v_n \equiv \tilde{w}_1({}_ia)[Xn]a\tilde{w}_2({}_ja)t[nX]taba\tilde{w}_3.$$

The extra condition required for this part is due to the fact that the occurrence vector of a in v_n is no longer identical to that of a in w. Once given this condition however we are still able to make the assumptions indicated at the start of this proof. The extra condition is still held in many commonly occurring cases: for example if a occurs more times in w than any other letter.

If there is a substitution θ such that $\theta(v_n)$ is a subword of w then $\theta(b) \equiv 1$ since there is no nontrivial subword between ${}_{(j+1)}a$ and ${}_{(j+2)}a$ in w yet in v_n the word b appears in this position. In this case $\theta(u_n) \equiv \theta(v_n)$. For the case where $\theta(u_n)$ is a subword of w, we may apply the arguments used in part (iv). □

Theorem 2.4.5 by no means captures all possible applications of Theorem 2.3.12.

For example in the word $w \equiv (ba)^n$ where $n > 2$, the vector $V_w(a)$ is not maximal (since $V_w(a) = V_w(b)$). Yet for every $n > 2$, $S(\{(ba)^n\})$ still satisfies the identity $u_n \equiv (bta)^{n-3}babt[Xn]ta[nX]a \approx (bta)^{n-3}bbat[Xn]ta[nX]a \equiv v_n$ since if θ is an assignment that does not assign a the value 1 and $\theta(u_n)$ is a subword of w then either $\theta(a) \equiv a$, $\theta(a) \equiv b$ or $\theta(a) = (ba)$ (these are the only subwords of w that occur as many times as the letter a does in u_n). If $\theta(a) \equiv ba$ then clearly $\theta(b) \equiv 1$ and $\theta(u_n) \equiv \theta(v_n)$. If $\theta(a) \equiv b$ then the first occurrence of a in u_n must be assigned the first occurrence of b in w. The first letter to appear in u_n is b and yet there is no letter left of the first occurrence of b in w. Therefore $\theta(b) \equiv 1$ and $\theta(u_n) \equiv \theta(v_n)$. The remaining case is when $\theta(a) \equiv a$ and then the proof becomes effectively the same as that of Theorem 2.4.5 part (iv). A similar argument applies when considering v_n. We have proved that the following is true.

EXAMPLE 2.4.6 *If $n > 2$ then $S(\{(ba)^n\})$ is NFB.*

Of course this also follows immediately from Theorem 2.0.10.

The arguments just used did not depend on the fact that $(ba)^n$ contained only two distinct letters, only on the fact that to the left of the first occurrence of b there was no proper subword occurring at least $n-1$ times (that is, the number of times that the letter b occurs in the identities used for Example 2.4.6). Thus we can deduce the following theorem.

THEOREM 2.4.7 *Let w be a word which has exactly two maximally occurring letters a and b with the first occurrence of b occurring in w before the first occurrence of a and with the property that every letter left of the first occurrence of b occurs fewer than $occ(a, w) - 1$ times. If w satisfies one of the conditions (i) to (vi) of Theorem 2.4.5 and the subwords described in the relevant part of Theorem 2.4.5 do not involve the first occurrence of a and of b in w then $S(\{w\})$ is NFB.*

DEFINITION 2.4.8 *An occurrence vector $V_w(u)$ of a subword u in a word w is said to be* super maximal *if the deletion of any one particular occurrence of u in w*

gives a new word v with the property that for any subword u' of w, $V_v(u) \leq V_w(u')$ only if $u' \equiv u$. Likewise $V_w(u)$ is super maximal in a set of words containing w if for every subword u' of a word $w' \in W$, $V_v(u) \leq V_{w'}(u')$ only if $u \equiv u'$ and $w \equiv w'$.

Clearly in this definition $V_v(u)$ can be obtained by subtracting the number 1 from one of the entries of $V_w(u)$ and deleting any zero entries from the resulting vector. A simple example of a super maximal occurrence vector is the occurrence vector of a letter in a word that has at least two extra occurrences in the word than any other letter.

We may now extend Lemma 2.4.4 as follows (the proof is similar to that of Lemma 2.4.4).

LEMMA 2.4.9 *Let W be a set of words, $w \in W$ be a word and u be a subword of w for which $V_w(u)$ is super maximal in W. If the occurrence vector $V_{w'}(u')$ of a subword u' in a second word w' (not necessarily in W) can be obtained by subtracting the number 1 from one of the entries of $V_w(u)$ and deleting any zero entries from the resulting sequence then for any substitution θ, $\theta(w')$ is a subword of a word in W only if $\theta(u') \equiv 1$ or both $\theta(w')$ is a subword of w and $\theta(u') \equiv u$.*

THEOREM 2.4.10 *Let W be a set of words and $w \in W$ be a word containing a letter a and a letter b such that $V_w(a)$ is super-maximal in W (the set W may of course be simply $\{w\}$ itself). Let p be any (possibly empty) word not containing the letter a and α_1, α_2, β_1 and β_2 be arbitrary positive integers. If w satisfies one of the following conditions or their reverse then $S(W)$ is NFB (in a similar way to before, we will assume that unless otherwise stated the given subwords may overlap but may not be contained within one another):*
(i) $apb^{\beta_1}aaa^{\alpha_1}b$ is a subword of w and $V_w(a)$ is the only occurrence vector of a subword in W that contains the occurrence vector of the letter a in the word obtained from w by replacing the given occurrence of $apb^{\beta_1}aaa^{\alpha_1}b$ by $apb^{\beta_1-1}aba^{\alpha_1}b$;
(ii) $baa^{\alpha_1}b$, apb and ba are subwords of w and the occurrence of ba in w does not

overlap with that of $baa^{\alpha_1}b$. *For example baababa, abaabba or baabbab is a subword of* w;

(iii) $apbb^{\beta_1}aa^{\alpha_1}b$ *is a subword of* w, *where* α *is maximal. For example abbaab is a subword of* w;

(iv) $b^{\beta_1}a^{\alpha_1}b^{\beta_2}a^{\alpha_2}pb$ *is a subword of* w, $\beta_1 > \beta_2$ *and* p *does not contain* b. *For example, bbabab is a subword of* w.

Proof: The proof of this theorem is similar to that of Theorem 2.4.5 except that the identities $u_n \approx v_n$ we construct in this case have only $occ(a,w) - 1$ occurrences of a (instead of $occ(a,w)$ occurrences). It is for this reason that we require $V_w(a)$ to be super maximal so that by Lemma 2.4.9 if θ is a substitution such that $\theta(u_n)$ (or $\theta(v_n)$) is a subword of a word in W, then either both $\theta(a) \equiv a$ and $\theta(u_n)$ (or $\theta(v_n)$) is a subword of w or $\theta(a) \equiv 1$, in which case $\theta(u_n) \equiv \theta(v_n)$.

First note that in every case in the theorem, w contains a subword of the form ab and another of the form ba (not intersecting). As in Theorem 2.4.5, this is all that is required to establish that $xytyx$ is an isoterm for $S(\{w\})$.

(i) Let $w \equiv w_1 apb^{\beta_1}({}_ia)aa^{\alpha_1}bw_2$.

Claim: $S(\{w\}) \models u_n \approx v_n$, where

$$u_n \equiv \tilde{w}_1 a t_1 [Xn] t_2 b^{\beta_1}({}_ia)a^{\alpha_1}[nX]b\tilde{w}_2$$

and

$$v_n \equiv \tilde{w}_1 a t_1 [Xn] t_2 b^{\beta_1 - 1}({}_ia)ba^{\alpha_1}[nX]b\tilde{w}_2$$

If θ is a substitution such that $\theta(u_n)$ is a subword of w then because $occ(a,u_n) = occ(a,w)-1$, θ must take the i^{th} occurrence of a in u_n to either the i^{th} or the $(i+1)^{th}$ occurrence of a in w (we will write this as $\theta({}_ia) \equiv ({}_ia)$ or ${}_{(i+1)}a$ in w). Now $\theta([nX])$ and $\theta(b)$ cannot contain a else we would have more than $occ(a,w)$ occurrences of a in $\theta(u_n)$. So in the first case (when $\theta({}_ia) \equiv ({}_ia)$) since the word a^{α_1+1} occurs immediately to the right of ${}_ia$ in w but the word $a^{\alpha_1}[nX]b$ occurs immediately to the right of ${}_ia$ in u_n, we must have $\theta([nX]b) \equiv 1$ or $\theta([nX]b)$ contains an occurrence

of a. If $\theta([nX]b)$ contains a then $\theta(u_n)$ contains more than $occ(a,w)$ occurrences of a which is not possible. In the second case (when $\theta({}_ia) \equiv ({}_{(i+1)}a)$), since there is the letter a immediately to the left of ${}_{(i+1)}a$ in w, the next letter left of ${}_ia$ in u_n not assigned the value 1 by θ, must be assigned a word ending in a. However, b cannot be assigned a word containing a. Consequently $\theta(b) \equiv 1$ and consequently $\theta(u_n) \equiv \theta(v_n)$.

Since $V_w(a)$ is the only occurrence vector of a letter in w that contains $V_{v_n}(a)$, we may assume as before that if θ is a substitution with the property that $\theta(v_n)$ is a subword of w then $\theta(a) \equiv a$. So θ must assign the i^{th} occurrence of a in v_n to either the i^{th} or the $(i+1)^{th}$ occurrence of a in w. In the first instance, ${}_{(i+1)}a$ lies immediately to the right of ${}_ia$ in w but in v_n, b lies immediately to the right of ${}_ia$. Since $\theta(b)$ does not contain a, $\theta(b)$ must be 1 and therefore $\theta(v_n) \equiv \theta(u_n)$. The second case follows in a similar way since immediately to the right of ${}_{(i+1)}a$ in w is the $(i+2)^{th}$ occurrence of a but b occurs to the right of ${}_ia$ in v_n.

(ii) In this case there are three subwords of w which we must consider. While any possible order of appearance of these subwords is allowed, as in Theorem 2.4.5 we need only consider one of these. Let $w \equiv w_1 b({}_ia)a^{\alpha_1} b w_2 a p b w_3 b a w_4$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1 b({}_ia)a^{\alpha_1 - 1}[Xn]b\tilde{w}_2 at[nX]tb\tilde{w}_3 ba\tilde{w}_4$$

and

$$v_n \equiv \tilde{w}_1 b({}_ia)a^{\alpha_1 - 1}[Xn]b\tilde{w}_2 at[nX]tb\tilde{w}_3 ab\tilde{w}_4.$$

Let θ be a substitution such that $\theta(u_n)$ (or $\theta(v_n)$) is a subword of w.

Case 1. $\theta({}_ia) \equiv ({}_ia)$ in w. Since to the right of ${}_ia$ in w we have a^{α_1}, $\theta(a^{\alpha_1 - 1}[Xn]b\tilde{w}_2)$ must be assigned a word starting with a^{α_1}. Since $\theta([Xn])$ and $\theta(b)$ cannot contain a (else there will be more than $occ(a,w)$ occurrences of a), they must be 1 and therefore $\theta(u_n) \equiv \theta(v_n)$.

Case 2. $\theta({}_ia) \equiv ({}_{(i+1)}a)$ in w. In this case, $\theta(\tilde{w}_1 b)$ must be assigned a word ending

in a, since to the left of ${}_{(i+1)}a$ in w there is the i^{th} occurrence of a. Since $\theta(b)$ doesn't contain a, it must be 1 and again $\theta(u_n) \equiv \theta(v_n)$.

(iii) Let $w \equiv w_1 apbb^{\beta_1}({}_ia)abw_2$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1 at[Xn]tb^{\beta_1}b({}_ia)[nX]b\tilde{w}_2$$

and

$$v_n \equiv \tilde{w}_1 at[Xn]tb^{\beta_1}({}_ia)b[nX]b\tilde{w}_2.$$

Let θ be a substitution such that $\theta(u_n)$ (or $\theta(v_n)$) is a subword of w. As before, this implies that $\theta(b)$ and $\theta([Xn])$ do not contain the letter a.

Case 1. $\theta({}_ia) \equiv ({}_ia)$ in w. If we are considering u_n, then $\theta(b) \equiv 1$ and $\theta(u_n) \equiv \theta(v_n)$ since there is an occurrence of a immediately to the right of ${}_ia$ in w but to the right of ${}_ia$ in u_n there is the word $[nX]b$. If we are considering v_n then $\theta(b)$ is 1 again since immediately to the right of ${}_ia$ in v_n is the letter b, but a occurs to the right of ${}_ia$ in w.

Case 2. $\theta({}_ia) \equiv ({}_{(i+1)}a)$ in w. In this case $\theta(u_n) \equiv \theta(v_n)$ since to the left of ${}_ia$ in both u_n and v_n is the letter b, but a occurs to the left of ${}_{(i+1)}a$ in w and therefore $\theta(b) \equiv 1$.

(iv) Let $w \equiv w_1 b^{\beta_1} a^{\alpha_1} b^{\beta_2}({}_ia)pbw_2$.

Claim: $S(\{w\}) \models u_n \approx v_n$ where

$$u_n \equiv \tilde{w}_1 b^{\beta_1} a^{\alpha_1 - 1}[Xn]b^{\beta_2}({}_{(i-1)}a)t[nX]tb\tilde{w}_2$$

and

$$v_n \equiv \tilde{w}_1 b^{\beta_1} a^{\alpha_1 - 1}[Xn]b^{\beta_2 - 1}({}_{(i-1)}a)bt[nX]tb\tilde{w}_2.$$

As usual, we will let θ be a substitution such that $\theta(u_n)$ (or $\theta(v_n)$) is a subword of w. This implies that $\theta(b)$ and $\theta([Xn])$ do not contain the letter a.

Case 1. $\theta({}_{(i-1)}a) \equiv {}_{(i-1)}\, a$. To the left of ${}_{(i-1)}a$ in w is the word $b^{\beta_1} a^{\alpha_1 - 1}$. To the left of ${}_{(i-1)}a$ in u_n is $a^{\alpha_1 - 1}[Xn]b^{\beta_2}$. Since $\beta_1 > \beta_2$, $\theta(a) \equiv a$ and $\theta(b)$ cannot be any

power of b greater than 1 (else there will be too many occurrences of b) we must have either $\theta(b) \equiv 1$ or $\theta(b) \equiv \theta([Xn]) \equiv b$. In the first case $\theta(u_n) \equiv \theta(v_n)$ and the second case never occurs since then $occ(b, \theta(u_n)) > occ(b, w)$. The case for v_n is similar since to the left of ${}_{(i-1)}a$ in v_n there is the word $a^{\alpha_1 - 1}[Xn]b^{\beta_2 - 1}$.

Case 2. $\theta({}_{(i-1)}a) \equiv_i a$. To the left of ${}_ia$ in w is the word ab^{β_2}. To the left of ${}_{(i-1)}a$ in u_n (and v_n) however we have the word $b^{\beta_1}[Xn]b^{\beta_2}$ (or $b^{\beta_1}[Xn]b^{\beta_2 - 1}$ respectively). Since $\beta_1 > \beta_2$ and neither $\theta(b)$ nor $\theta([Xn])$ contain a, we must have $\theta(b) \equiv 1$. The proof is complete. □

The previous two theorems followed from Theorem 2.3.12. We now present an analogous theorem using Theorem 2.3.14.

THEOREM 2.4.11 *Let W be a set of words and $w \in W$ be a word containing letters a, b, c for which $V_w(a)$ is maximal in W and let u and v be any (possibly empty) words with $a, b \notin c(u)$ and $a, c \notin c(v)$ (the set W may of course be simply $\{w\}$ itself). If w has one of the following properties (or their reverse) then $S(W)$ is not finitely based:*

(i) bac and aucabva are non overlapping subwords of w;

(ii) bacva and aucab are non overlapping subwords of w;

(iii) bac and avbacua are non overlapping subwords of w;

(iv) avbac and avbac are non overlapping subwords of w.

Proof: Parts (i) and (ii) are obtained by an application of part (i) of Theorem 2.3.14 with $u_1 \equiv b$ and $u_2 \equiv c$. Since the two proofs are almost identical we will only prove part (i) here. Likewise parts (iii) and (iv) follow in a very similar manner from part (ii) of Theorem 2.3.14 and so will also not be proved. Since $V_w(a)$ is maximal in w, Lemma 2.4.4 implies that if θ is a substitution so that $\theta(r_n)$ (or $\theta(r'_n)$) is a subword of w, then either $\theta(a) \equiv a$ or $\theta(a) \equiv 1$. As in the previous two theorems we will not concern ourselves with the order of appearance of the given subwords in w.

(i) Let $w \equiv w_1 b({}_i a) c w_2 a u c({}_j a) b v a w_3$. We will show that $S(\{w\})$ satisfies the identity

$$\ddot{w}_1 a t_1 [Xn]({}_i a) t_2 a \ddot{w}_2 t [nX]({}_j a) t \ddot{w}_3 \approx \ddot{w}_1 a t_1 ({}_i a)[Xn] t_2 a \ddot{w}_2 t ({}_j a)[nX] t \ddot{w}_3.$$

Firstly if $\theta([Xn]) \equiv 1$ or $\theta(a) \equiv 1$ then $\theta(r_n) \equiv \theta(r'_n)$. Left of ${}_i a$ in w is the letter b. So if $\theta(r_n)$ is a subword of w and $\theta([Xn]) \not\equiv 1$, then $\theta([Xn])$ contains b. But then $\theta([nX])$ contains b and so contained in the word between ${}_{(j-1)}a$ and ${}_j a$ in $\theta(r_n)$ is a letter b. However between ${}_{(j-1)}a$ and ${}_j a$ in w there is no letter b, contradicting the assumption that $\theta(r_n)$ was a subword of w. The case when $\theta(r'_n)$ is a subword of w follows by symmetry. □

The following corollary is a dual version of Corollary 2.2.6 and follows immediately from the proofs of Theorems 2.4.5, 2.4.10 and 2.4.11.

COROLLARY 2.4.12 *Let w be a word satisfying the conditions of Theorem 2.4.5, 2.4.10 or 2.4.11 and let $r = max\{occ(x, w) : x \in c(w) \backslash \{a\}\}$. If W is a set of r-limited words then $S(W \cup \{w\})$ is NFB.*

It is clear that the word w in this corollary can be taken from a two letter alphabet. Combining Corollaries 2.2.6 and 2.4.12 we obtain

COROLLARY 2.4.13 *If W is a set of words then there are sets of words $W = V_0, V_1, V_2, \ldots$ with $|c(V_i)| = max(2, |c(W)|)$ and $V_i \subset V_{i+1}$ for $i \geq 0$ so that $S(V_{2j})$ is FB and $S(V_{2j+1})$ is NFB for every $j > 0$.*

A further result is the following.

THEOREM 2.4.14 *If* **S** *is a k-nilpotent monoid then* **S** *is a subsemigroup of a NFB $max(5, (k+3))$-nilpotent monoid which is finite if* **S** *is finite.*

Proof: We will assume that $k \geq 2$ and show that S is a subsemigroup of a $k+3$ nilpotent monoid. Let k' be the smallest integer such that $2k' > k$ for some number k. Consider the monoid $S(\{(ba)^{k'}\})$. This is certainly $(k+3)$-nilpotent since the

length of $(ba)^{k'}$ is either $k+1$ or $k+2$ and in both cases $S(\{(ba)^{k'}\})$ is $(k+3)$-nilpotent. Using the same construction as for Corollary 2.2.5 we arrive at a $(k+3)$-nilpotent monoid $\mathbf{T}$ containing both $\mathbf{S}$ and $S(\{(ba)^{k'}\})$ as subsemigroups. To show that $\mathbf{T}$ is NFB we now use the identities $u_n \approx v_n$ of Example 2.4.6 (if $k' > 2$) or the identities

$$\{[X(2n)]t[\mathcal{X}(2n)] \approx [\mathcal{X}(2n)]t[X(2n)] : n \in \mathbb{N}\}$$

of Lemma 2.3.3 (if $k' = 2$). For the remainder of this proof it will be convenient to denote this last set of identities by Σ.

If $k' > 2$ then the semigroup $S(\{(ab)^{k'}\})$ satisfies $u_n \approx v_n$ and since the only unstable pair in these identities is (a, b) and $|u_n(a, b)| = |v_n(a, b)| \geq k$, so $\mathbf{S}$ must also satisfy $u_n \approx v_n$ (since a word w of length k takes the value 0 on $\mathbf{S}$ unless 1 is assigned to at least one letter in $c(w)$). Therefore $\mathbf{T}$ satisfies $u_n \approx v_n$. If $k' = 2$ then using effectively the same arguments as above we see that $\mathbf{S}$, $S(\{abab\})$ and therefore $\mathbf{T}$ all satisfy the identities Σ of Lemma 2.3.3. If $k' > 2$ then $\mathbf{T}$ is NFB by Theorem 2.4.5 part (iv). If $k' = 2$ then $\mathbf{T}$ is NFB by Lemma 2.3.3. □

An immediate corollary of this is

COROLLARY 2.4.15 *The pseudovariety generated by the class of finite NFB nilpotent monoids contains all finite nilpotent semigroups and all nilpotent monoids.*

This corollary and Corollary 2.2.5 show that both the class of finite FB nilpotent monoids and the class of finite NFB nilpotent monoids generate the same pseudovariety as that generated by the class of all finite nilpotent monoids. H. Straubing [89] has shown that this pseudovariety is exactly the class of finite aperiodic semigroups with central idempotents (that is, finite aperiodic semigroups which satisfy $e^2 = e \rightarrow ex = xe$).

COROLLARY 2.4.16 *If* $\mathbf{S}$ *is a finite aperiodic semigroup with central idempotents then there are sets of words* $V_1, V_2, \ldots$ *with* $|c(V_i)| = 2$ *and* $V_i \subset V_{i+1}$ *for* $i \geq 0$ *so that for every* $j > 0$, $\mathcal{V}(S(V_{2j}))$ *is FB,* $\mathcal{V}(S(V_{2j+1}))$ *is NFB and* $\mathbf{S} \in \mathcal{V}(S(V_1)) \subset \mathcal{V}(S(V_2)) \subset \ldots$.

As an example of the power of the theorems in this section we briefly examine an amusing though useless "application" of the results to genetics.

EXAMPLE 2.4.17 *In genetics, the base sequence for a DNA molecule can be thought of as a long word, w, in the alphabet $\{a, c, g, t\}$ (see Figure 41 of [26] for an example of a very short base sequence). The molecules corresponding to these letters are called* bases. *In such a large word, it is extremely likely to find for any pair of letters in this alphabet the subwords required for applications of Theorem 2.4.5 or 2.4.7. Now for a given strand of DNA the word w obviously contains many occurrences of each of the letters a, c, g and t, and it would appear to be unlikely that two of these letters would occur exactly the same number of times. Further evidence for this claim can be found in the results of [35] for example where it is shown that only 32 % of the base sequence for the DNA of the Antarctic krill* Euphausia superba *is a g or a c (similar results hold for most other organisms as well). Thus the letters a and t (or at least one of these letters) occur a significantly greater amount of the time than do g or c. If, in a particular strand of DNA, one of the bases a or t occurs more times than any other base then Theorem 2.4.5 implies that the discrete syntactic monoid of the corresponding base sequence is NFB (given that the appropriate subwords for the application of this theorem are plentiful). Using Theorem 2.4.7 we obtain the same result if the two bases a and t occur the same number of times in a particular strand (though this event would seem unlikely).*

While this example may not be of interest to geneticists, it does illustrate the ability of Theorem 2.4.5 (and Theorem 2.4.7) to apply to long and complicated words.

COROLLARY 2.4.18 *Every word w is a subword of a word w' whose length is no more than 4 letters longer than w and such that $S(\{w'\})$ is NFB. If $|c(w)| > 1$ then w' can be chosen such that $c(w') = c(w)$.*

Proof: If $|c(w)| = 0$ then w is the empty word and it follows from Theorem 2.4.2 that the shortest word containing w whose discrete syntactic monoid is not finitely based is the word $abab$ or $abba$.

If $|c(w)| = 1$ then w is of the form a^k for some k. In this case we may choose w' to be the word $a^k bab$ for some new letter b. Using exactly the same argument as for Example 2.3.4 it follows that $S(\{w'\})$ is NFB by Lemma 2.3.3. Now assume $|c(w)| > 1$.

Case 1. w ends with a letter a that occurs a maximal number of times in w and there is at least one letter b occurring in w fewer times than a. In this case we may take w' to be the word $wbaba$ and apply Theorem 2.4.5 part (iv).

Case 2. Every letter of w occurs an equal number of times. Let b be the last letter in w and a be the next letter left of this that is different to b. So $w \equiv w_1 ab^\beta$ for some $\beta > 0$. Thus we may take w' to be the word $w_1 ab^\beta aaab$ and apply Theorem 2.4.10 part (i).

Case 3. w ends with a letter, b say, not occurring a maximal number of times in w. Let a be the closest letter to the right end of w that does occur a maximal number of times. Then we may choose w' as the word $waaab$ and apply Theorem 2.4.10 part (i). □

EXAMPLE 2.4.19 *Consider the monoid $S(\{abcbadefgef\})$. This semigroup is in fact FB (this will be shown later; see page 91) but Corollary 2.4.18 implies that the semigroup $S(\{abcbadefgefgfgf\})$ is NFB. Note also that Theorem 2.4.5 part (iv) implies that the semigroup $S(\{fefabcbadefgef\})$ is NFB.*

2.5 On the Finite Basis Problem for almost all discrete syntactic monoids of k element languages in fixed finite alphabets

The conditions contained in Theorems 2.4.5, 2.4.10, and 2.4.11 are very general. After a little experimentation it becomes clear that for sufficiently long words in any fixed finite alphabet the likelihood of one of these theorems applying is very high (this was exploited in Example 2.4.17). In this section we investigate this apparent property and show that these theorems in fact apply to "almost all" words w (and in some sense, sets of words W) in a fixed alphabet. First we formally define the notion of "almost all".

Recall that the length of a set of words W is $max\{|w| : w \in W\}$. Fix an alphabet A and let $W_{(1,n,k)}$ be the set of all k element length n sets of words from the free monoid A^* and $N_{(1,n,k)}$ be the number of elements of $W_{(1,n,k)}$ (each of these elements are k element, length n sets of words from A^*). Now let P be a property and $W_{(P,n,k)}$ be the set of all k element length n sets of words from the free monoid A^* which have the property P. Following the above notation, we will use $N_{(P,n,k)}$ to denote the number of elements of $W_{(P,n,k)}$.

Note that if one word in a k element length n sets of words W is a proper subword of another word in W, then $S(W)$ is identical to the discrete syntactic monoid of a language with fewer elements.

DEFINITION 2.5.1 *For a given positive integer k and a finite alphabet A, a property P holds for almost all k element sets of words in A^* if $N_{(P,n,k)}/N_{(1,n,k)} \to 1$ as $n \to \infty$ (or equivalently if $(N_{(1,n,k)} - N_{(P,n,k)})/N_{(1,n,k)} \to 0$ as $n \to \infty$).*

In general for sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ we will write $s_n \sim t_n$ if

$$\lim_{n\to\infty} s_n/t_n \to 1.$$

That is, given a fixed finite alphabet A, a property P holds for almost all k element sets of words in A if and only if $N_{(P,n,k)} \sim N_{(1,n,k)}$. When $k = 1$ in the above, we are considering one element sets of words and the length of such set is simply the length of the unique word it contains. In this case we will abbreviate $W_{(P,n,1)}$ and $N_{(P,n,1)}$ to $W_{(P,n)}$ and $N_{(P,n)}$ respectively and say that the property P holds for almost all words if $N_{(P,n)} \sim N_{(1,n)}$.

We now establish some basic facts concerning the combining of properties that hold for almost all k element sets of words of a finite alphabet. It is easy to verify that the relation $\sim$ is an equivalence since for any properties P, Q, R, we have: $N_{(P,n,k)} \sim N_{(P,n,k)}$; $N_{(P,n,k)} \sim N_{(Q,n,k)} \Rightarrow N_{(Q,n,k)} \sim N_{(P,n,k)}$; and if both $N_{(P,n,k)} \sim N_{(Q,n,k)}$ and $N_{(Q,n,k)} \sim N_{(R,n,k)}$ then $N_{(P,n,k)} \sim N_{(R,n,k)}$. A further important property of the relation $\sim$ is given in the following lemma. Here if P and Q are properties, then $P \cap Q$ is the property of having both the properties P and Q.

LEMMA 2.5.2 *For any fixed finite alphabet, if* $N_{(P,n,k)} \sim N_{(1,n,k)}$ *and* $N_{(Q,n,k)} \sim N_{(1,n,k)}$ *then* $N_{(P\cap Q,n,k)} \sim N_{(1,n,k)}$.

Proof: We want to show that $N_{(P\cap Q,n,k)}/N_{(1,n,k)} \to 1$ as $n \to \infty$. Now

$$W_{(1,n,k)} \backslash (W_{(P,n,k)} \cap W_{(Q,n,k)}) =$$
$$(W_{(P,n,k)} \backslash W_{(Q,n,k)}) \cup (W_{(Q,n,k)} \backslash W_{(P,n,k)}) \cup (W_{(1,n,k)} \backslash (W_{(P,n,k)} \cup W_{(Q,n,k)})).$$

But $W_{(P,n,k)} \backslash W_{(Q,n,k)} \subseteq W_{(1,n,k)} \backslash W_{(Q,n,k)}$, $W_{(Q,n,k)} \backslash W_{(P,n,k)} \subseteq W_{(1,n,k)} \backslash W_{(P,n,k)}$ and

$$(W_{(1,n,k)} \backslash (W_{(P,n,k)} \cup W_{(Q,n,k)})) \subseteq W_{(1,n,k)} \backslash W_{(P,n,k)}.$$

So therefore

$$\frac{|W_{(1,n,k)} \backslash (W_{(P,n,k)} \cap W_{(Q,n,k)})|}{N_{(1,n,k)}} \leq 2\frac{|W_{(1,n,k)} \backslash W_{(P,n,k)}|}{N_{(1,n,k)}} + \frac{|W_{(1,n,k)} \backslash W_{(Q,n,k)}|}{N_{(1,n,k)}}$$

which tends toward 0 as n tends to infinity since both

$$\frac{|W_{(1,n,k)} \backslash W_{(P,n,k)}|}{N_{(1,n,k)}} \text{ and } \frac{|W_{(1,n,k)} \backslash W_{(Q,n,k)}|}{N_{(1,n,k)}}$$

tend toward 0 as n tends to infinity. Therefore $\frac{N_{(1,n,k)}-N_{(P\cap Q,n,k)}}{N_{(1,n,k)}} \to 0$ as $n \to \infty$ and $N_{(P\cap Q,n,k)} \sim N_{(1,n,k)}$ as required. □

The proportion of k element, length n sets of words that have a property P is exactly the probability of selecting a set of words at random from $W_{(1,n,k)}$ that has the property P. For this reason it is convenient to interpret problems concerning the ratio $N_{(P,n,k)}/N_{(1,n,k)}$ in terms of probability.

LEMMA 2.5.3 *If w is a word in a finite alphabet A then almost all words in A^* have w as a subword.*

Proof: Let $|A| = r$ and the length of w be m. For any n there are r^n words of length n. Therefore the likelihood of a randomly chosen word of length m being the word w is exactly $1/r^m$. Any word w' of length n can be partitioned into $[n/m]$ (where $[n/m]$ denotes the integer part of n/m) subwords of length m along with a remaining subword of length less than m. If w' does not contain w as a subword, then it is necessary that each of these partitions is not the word w. Thus for a word w' of length n the likelihood that that w' does not contain w as a subword is less than or equal to $(1 - 1/r^m)^{[n/m]}$. Since $1 - 1/r^m = (r^m - 1)/r^m < 1$, it must be that $((r^m - 1)/r^m)^{[n/m]} \to 0$ as $n \to \infty$. That is, almost all words in A^* have w as a subword. □

Ultimately we want to show that almost all words and almost all k element sets of words in a fixed finite alphabet have discrete syntactic monoids that are NFB. In order to apply the most general theorems of the previous section we need to show that one can find a maximal occurrence vector for a letter in almost all words in a fixed finite alphabet (and a maximal occurrence vector in almost all k element sets of words). As discussed earlier (see page 58), one of the simplest ways that this can happen is if almost all words have a unique maximally occurring letter. The next few lemmas establish this fact.

LEMMA 2.5.4 *Let $A = \{a_1, \ldots, a_r\}$ be a fixed finite alphabet of $r \geq 2$ distinct letters. Let a word w in A^+ have the property P if it contains no letter a so that $occ(a, w) > occ(x, w)$ for all $x \in c(w) \backslash \{a\}$, that is, that there is no unique maximally occurring letter in w. Then*

$$N_{(P,n)}/N_{(1,n)} \leq \binom{r}{2} \sum_{n/r \leq m \leq n/2}^{n} \binom{2m}{m} (1/2)^{2m} \binom{n}{2m} (2/r)^{2m} (1 - 2/r)^{n-2m}$$

Proof: Let X_i be a random variable corresponding to the number of occurrences of the letter a_i in a word of length n. Each successive letter appearing in the word can be thought of as the outcome of a Bernoulli trial, with the appearance of the letter a_i (which occurs with probability $1/r$) considered a success and the appearance of any other letter considered a failure. Evidently X_i is binomially distributed and the probability of X_i taking a particular value $x \leq n$ is given by $\binom{n}{x}(1/r)^x(1-1/r)^{n-x}$ (information regarding the Binomial distribution $Bi(n, \theta)$ can be found in many books concerning probability or statistics; see [87] for example). For distinct numbers $i, j \leq n$, the variables X_i and X_j are not independent since the number of occurrences, say m, of the letter a_i in a given word of length n reduces the potential number of occurrences of the letter a_j to $n - m$. However the distribution of the sum $X_i + X_j$ is easily seen to be $Bi(n, 2/r)$ and given a particular value of $X_i + X_j$, say k, the probability that X_i takes some value m (necessarily less than or equal to k) is $\binom{k}{m}(1/2)^m(1 - 1/2)^{k-m} = \binom{k}{m}(1/2)^k$.

Let E be the event that there is no unique maximally occurring letter in a word w, that is that w has the property P. Clearly a letter that occurs less than n/r times in a word of length n in an r letter alphabet cannot be a maximally occurring letter. Therefore

$$E \subseteq \{X_i = X_j \geq n/r, \text{ for some } i \neq j\}$$

and so

$$\begin{aligned} N_{(P,n)}/N_{(1,n)} =& Pr(E) \\ \leq& Pr(X_i = X_j \geq n/r, \text{ for some } i \neq j) \\ =& Pr(\bigcup_{i<j} \{X_i = X_j \geq n/r\}) \\ \leq& \sum_{i<j} Pr(X_i = X_j \geq n/r) \\ =& \binom{r}{2} Pr(X_1 = X_2 \geq n/r). \end{aligned}$$

Now let $S = X_1 + X_2$. The distribution of S is $Bi(n, 2/r)$ and so

$$\begin{aligned} & Pr(X_1 = X_2 \geq n/r) \\ &= \sum_{k=0}^{n} Pr(X_1 = X_2 \geq n/r | S = k) Pr(S = k) \\ &= \sum_{0 \leq m \leq n/2} Pr(X_1 = m \geq n/r | S = 2m) Pr(S = 2m) \\ &= \sum_{n/r \leq m \leq n/2} Pr(X_1 = m | S = 2m) Pr(S = 2m) \\ &= \sum_{n/r \leq m \leq n/2} \left(\binom{2m}{m} (1/2)^{2m} \right) \left(\binom{n}{2m} (2/r)^{2m} (1 - 2/r)^{n-2m} \right) \end{aligned}$$

as required. □

We now want to show that the bound for $N_{P,n}/N_{1,n}$ obtained in Lemma 2.5.4 tends toward 0 as n tends toward infinity. The following lemma proved by B. M. Brown establishes this fact. Since this lemma is unpublished we present its proof here for the sake of completeness.

LEMMA 2.5.5 *(B. M. Brown, private communication)*

$$\sum_{n/r \leq n/2} \binom{2m}{m} (1/2)^{2m} \binom{n}{2m} (2/r)^{2m} (1 - 2/r)^{n-2m} \leq \frac{1}{2} \sqrt{\frac{r}{n\pi}}.$$

Proof: We first use Legendre's duplication formula (see page 5 of [16] for example)

$$\sqrt{\pi} \Gamma(2z) = 2^{2z-1} \Gamma(z) \Gamma(z + 1/2),$$

where $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt$ is the Gamma function. This implies

$$\begin{aligned}\binom{2m}{m} = \frac{(2m)!}{(m!)^2} &= \frac{\Gamma(2m+1)}{\Gamma^2(m+1)} \\ &= \frac{2m\Gamma(2m)}{m^2\Gamma^2(m)} \\ &= \frac{2}{m}\frac{2^{2m-1}\Gamma(m)\Gamma(m+1/2)}{\sqrt{\pi}\Gamma^2(m)} \\ &= \frac{2^{2m}}{m\sqrt{\pi}}\frac{\Gamma(m+1/2)}{\Gamma(m)}.\end{aligned}$$

This means that the expression on the left of the lemma can be reduced to

$$\sum_{n/r \leq n/2} \binom{n}{2m}(2/r)^{2m}(1-2/r)^{n-2m}\frac{1}{m\sqrt{\pi}}\frac{\Gamma(m+1/2)}{\Gamma(m)}$$

To complete the proof we now need to examine the term $\frac{\Gamma(m+1/2)}{\Gamma(m)}$.

LEMMA 2.5.6 $\frac{\Gamma(m+1/2)}{\Gamma(m)} \leq \sqrt{m}$.

Proof: The proof suggested by B. M. Brown used the product form of the Gamma function (see page 1 of [16] for example). Instead we use a simpler argument based on Stirling's formula. It is well known that Stirling's formula for factorials can be extended to the Gamma function; indeed it follows from one proof of Stirling's formula that

$$\Gamma(x) = \sqrt{2\pi}x^{x-1/2}e^{-x+\mu(x)}$$

where $\mu(x)$ decreases monotonically toward a limit of 0 as x tends toward infinity (see Chapter 3 of [1]; two alternative proofs of Stirling's formula may also be found

in [7]). Therefore $e^{\mu(x+1/2)}/e^{\mu(x)} < 1$ and so

$$\begin{aligned}\frac{\Gamma(m+1/2)}{\Gamma(m)} \leq & \frac{(m+1/2)^m}{e^{m+1/2}} \frac{e^m}{(m^{m-1/2})} \\ = & \frac{(m+1/2)^m}{m^m} \frac{\sqrt{m}}{\sqrt{e}} \\ = & \left(\frac{(m+1/2)}{m}\right)^m \frac{\sqrt{m}}{\sqrt{e}} \\ \leq & \sqrt{e} \frac{\sqrt{m}}{\sqrt{e}} \\ = & \sqrt{m}\end{aligned}$$

as required (note that it also follows from this proof that $\Gamma(m+1/2)/\Gamma(m) \sim \sqrt{m}$).□

We may now complete the proof of Lemma 2.5.5. Let $g_m = \frac{\Gamma(m+1/2)}{m\sqrt{\pi}\Gamma(m)} \leq 1/\sqrt{m\pi}$ and let

$$h_j = \begin{cases} g_{j/2} & \text{if } j \text{ is even and } j \geq (2n)/r, \\ 0 & \text{otherwise.} \end{cases}$$

Now if the distribution of X is $Bi(n, 2/r)$ and $Y = f(X)$ is a random variable that depends on X then the expected value $E(Y)$ is

$$\sum_{k=0}^{n} P(X = k) f(k) = \sum_{k=0}^{n} \binom{n}{i} (2/r)^k (1 - 2/r)^{n-k} f(k).$$

So the expression

$$\sum_{n/r \leq m \leq n/2} \binom{n}{2m} (2/r)^{2m} (1 - 2/r)^{n-2m} \frac{1}{m\sqrt{\pi}} \frac{\Gamma(m+1/2)}{\Gamma(m)}$$

is the expected value $E(h_X)$ where the distribution of X is $Bi(n, 2/r)$. But

$$X = (2n)/r + z_n \sqrt{n(2/r)(1 - 2/r)}$$

where by the Central Limit Theorem for binomially distributed random variables the distribution of z_n tends toward $N(0, 1)$ (the standard normal distribution) as n

tends toward infinity. (Here we use the notation z_n instead of the usual Z_n to avoid confusion between the Zimin words Z_n of Theorem 1.1.1.)

The value of h_j is alternately $g_{j/2}$ or 0 as j is even or odd. It follows from a simple examination of binomial probabilities for even integer values that as $E(h_X) \sim 1/2E(g_{X/2})\chi(X/2 \geq n/r)$ (where for a condition C, the value of $\chi(C)$ is 1 if the condition C is true and 0 otherwise). Now by Lemma 2.5.6,

$$g_{\{X/2\}} = g_{\{n/r+z_n\sqrt{(n/2r)(1-2/r)}\}} \leq \frac{1}{\sqrt{\pi}}\left(n/r + z_n\sqrt{\frac{n}{2r}\left(1-\frac{2}{r}\right)}\right)^{-1/2}$$

when $X/2 \geq n/r$. But

$$\left(n/r + z_n\sqrt{\frac{n}{2r}\left(1-\frac{2}{r}\right)}\right)^{-1/2} \leq \sqrt{r/n}$$

when $z_n \geq 0$ (or equivalently, when $X/2 \geq n/r$). So $1/2E(g_{X/2}) \leq \frac{1}{2}\sqrt{\frac{r}{n\pi}}$ and therefore $E(h_X) \leq \frac{1}{2}\sqrt{\frac{r}{n\pi}}$ as required. □

Combining Lemma 2.5.4 and Lemma 2.5.5 we have the following lemma.

LEMMA 2.5.7 *Almost all words in a fixed finite alphabet have the property that there is a unique maximally occurring letter.*

Proof: This is because the property P' that a word has a unique maximally occurring letter is the compliment of the property P in Lemma 2.5.4. Since by Lemma 2.5.5, $\lim_{n\to\infty} N_{(P,n)}/N_{(1,n)} = 0$, it must be the case that $\lim_{n\to\infty} N_{(P',n)}/N_{(1,n)} = 1$ as required. □

To generalise Lemma 2.5.7 to k element sets of words it is necessary to obtain variations of Lemmas 2.5.4 and 2.5.5. First note that a k element, length n set of words W in a finite alphabet A can be constructed by first selecting a word of length n from A^* and then selecting $k-1$ distinct words from the remaining words in A^* that have length at most n. In the lemma to follow it is convenient to relax the condition that these words must all be distinct. In this case it is possible that the set of words constructed actually has fewer than k elements and so is not of the

desired form. However if $|A| = r$, then the total number of words of length at most n in A^* is easily seen to be $r^n + r^{n-1} + \ldots + r$. Thus in the process of selecting k (not necessarily distinct) words in a length n set of words W, the likelihood of a word being selected twice is proportional to $1/(r^n + r^{n-1} + \ldots + r) \leq 1/r^n$. That is, almost all selections of k words of length at most n (including at least one of length n) have no repeats.

LEMMA 2.5.8 *Let $A = \{a_1, \ldots, a_r\}$ be a fixed finite alphabet of $r \geq 2$ distinct letters. Let P be the property of a k element set $W = \{w_1, \ldots, w_k\}$ of words of maximum length n and in the alphabet A that for every letter $a \in A$ occurring at least n/r times in a word w in W, there is a distinct word $v \in W$ containing a (possibly identical) letter b so that $occ(a, w) = occ(b, v)$. Then*

$$\frac{N_{(P,n,k)}}{N_{(1,n,k)}} \leq \binom{rk}{2} \sum_{n/r \leq m \leq n} \left(\binom{n}{k} (1/r)^k (1 - 1/(r))^{n-k} \right)^2 .$$

Proof: Note that allowing for repeated words in W actually allows for extra ways in which a letter can occur the same number of times in different words. Thus the true value for $\frac{N_{(P,n,k)}}{N_{(1,n,k)}}$ is likely to be smaller than that obtained in this lemma.

Since the longest word in a length n set of words $W = \{w_1, \ldots, w_k\}$ has length n we may assume without loss of generality that the length of w_1 is exactly n, although for some $i > 1$ it is possible that $|w_i| < n$. Let $X_{i,j}$ be the random variable corresponding to the number of occurrences of the letter a_j in the word w_i. As in Lemma 2.5.4 we will only be concerned with the situation when $X_{i,j} \geq |w_i|/r$. If $|w_i| < n$ then the expected value $E(X_{i,j})$ is $|w_i|/r < n/r$ and so the probability $Pr(X_{i,j} = m \geq n/r)$ is less than $Pr(X_{1,j} = m \geq n/r)$. Therefore it suffices to prove the lemma in the case when every word has length n.

Let E be the event that a k element set of words from A^* has the property P, that is, the event

$$\{X_{i,j} = X_{i',j'} \geq n/r, \text{ for some} i' \neq i\}$$

and so

$$\begin{aligned}\frac{N_{(P,n,k)}}{N_{(1,n,k)}} &= Pr(E)\\ &= Pr(\{X_{i,j} = X_{i',j'} \geq n/r, \text{ for some } i' \neq i\})\\ &= Pr\left(\bigcup_{i<i'}\{X_{i,j} = X_{i',j'} \geq n/r\}\right)\\ &\leq \sum_{i<i'} Pr(X_{i,j} = X_{i',j'} \geq n/r)\\ &= \binom{rk}{2} Pr(X_{1,1} = X_{2,1} \geq n/r).\end{aligned}$$

Now the event $\{X_{1,1} = X_{2,1} \geq n/r\}$ is exactly the event $\cup_{n/r\leq i\leq n}\{X_{1,1} = X_{2,1} = i\}$. Since the probability $P(X_{1,1} = i) = \binom{n}{i}(1/r)^i(1-1/r)^{n-i}$ we must have

$$Pr(X_{1,1} = X_{2,1} \geq n/r) \leq \sum_{n/r\leq i\leq n}\left(\binom{n}{i}(1/r)^i(1-1/r)^{n-i}\right)^2.$$

Note that the word w_2 is distinct from the word w_1 and so the probability that $X_{2,1}$ takes on a value i given that $X_{1,1} = i$ is actually less than $\binom{n}{i}(1/r)^i(1-1/r)^{n-i}$. □

When $r = 6$, for example, the probability $\sum_{n/r\leq k\leq n}\left(\binom{n}{k}(1/r)^k(1-1/r)^{n-k}\right)^2$ is exactly the likelihood of rolling two fair die n times and obtaining exactly the same number of 1's from the first dice as from the second and having this number greater than or equal to $n/6$. Intuitively, one might expect that as n increases toward infinity, the value of this probability decreases toward zero; indeed this is what we now prove.

LEMMA 2.5.9

$$\sum_{n/r\leq k\leq n}\left(\binom{n}{k}(1/r)^k(1-1/r)^{n-k}\right)^2 \sim r/\sqrt{n\pi(r-1)}$$

Proof: We first note that by the Central Limit Theorem the distribution of both $X_{1,1}$ and $X_{1,2}$ is increasingly well approximated by $N(\mu, \sigma^2)$ where μ is the expected value

$E(X_{1,1}) = n/r$ and σ^2 is the variance $var(X_{1,1}) = var(X_{2,1}) = (n/r)(1-1/r) = n(r-1)/r^2$. The probability density function for this distribution is

$$f_{X_{1,1}}(x) = \frac{1}{\sigma\sqrt{2\pi}} Exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

Because $f_{X_{1,1}}(x)$ is the probability density function of a random variable with expected value μ, we must have

$$\int_{\mu}^{\infty} f_{X_{1,1}}(x)dx = 1/2$$

and so

$$\int_{\mu}^{\infty} Exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \sigma\sqrt{\pi/2}.$$

Therefore

$$\begin{aligned}
&\sum_{n/r\leq k\leq n} \left(\binom{n}{k}(1/r)^k(1-1/r)^{n-k}\right)^2 \\
&\sim \int_{\mu}^{\infty} \left(\frac{1}{\sigma\sqrt{2\pi}} Exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\right)^2 dx \\
&= \frac{1}{\sigma^2 2\pi}\int_{\mu}^{\infty} Exp\left(-\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\
&= \frac{1}{\sigma^2 2\pi}\int_{\mu}^{\infty} Exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma/\sqrt{2}}\right)^2\right) dx \\
&= \frac{1}{\sigma^2 2\pi}\left((\sigma/\sqrt{2})(\sqrt{2\pi})\right) \\
&= \frac{1}{\sigma\sqrt{\pi}} \\
&= r/\sqrt{n\pi(r-1)}
\end{aligned}$$

as required. □

We can now prove the following lemma.

LEMMA 2.5.10 *For any $k > 0$, almost all k element sets of words in a fixed finite alphabet have the property that there is a unique, maximally occurring letter.*

Proof: It follows from Lemmas 2.5.8 and 2.5.9 that almost all k element sets of words in a fixed finite alphabet have the property that there is a unique word whose maximally occurring letters occur more times than in any other word in the set. By Lemma 2.5.7, however, this word almost always has a unique maximally occurring letter. Therefore almost all k element sets of words in a fixed finite alphabet have the desired property. □

We now have all the results needed to prove the main theorem of this section. If A is an alphabet consisting of a single letter then for all words w and all sets of words W from A^*, the monoids $S(\{w\})$ and $S(W)$ are easily seen to be FB. We now show that for $|A| > 1$ the opposite is nearly true.

THEOREM 2.5.11 *Let A be a finite alphabet with $|A| > 1$ and k be a fixed positive integer. Then for almost all k element sets of words $W \subset A^*$, $S(W)$ is not finitely based.*

Proof: Combining Lemmas 2.5.2, 2.5.3 and 2.5.10 it follows that almost all k element sets of words W in a fixed finite alphabet contain a word w with a letter a occurring more times than any other letter in $c(W)$ and that w contains the word, say, *abbbab* as a subword for some letter $b \in c(w)$. The occurrence vector of a in w is maximal in W so by Theorem 2.4.5, $S(W)$ is almost always not finitely based. □

COROLLARY 2.5.12 *For any fixed positive letter k, almost all discrete syntactic monoids of k element languages from a fixed finite alphabet A are not finitely based.*

Proof: If two discrete syntactic monoids of k element sets of words W_1 and W_2 in a finite alphabet A are isomorphic then the sets W_1 and W_2 must have the same length, say n. There is exactly one minimal generating set for each of $S(W_1)$ and $S(W_2)$ and these are $c(W_1)$ and $c(W_2)$ respectively. Therefore we may assume that $c(W_1) = c(W_2)$ and that any isomorphism $\iota : S(W_1) \to S(W_2)$ must restrict to a permutation of $c(W_1) = c(W_2) \subseteq A$. Clearly this permutation along with the

multiplication of $S(W_1)$ determines the multiplication on $S(W_2)$. There are at most $|A|!$ permutations of $c(W_1)$ and therefore there are at most $|A|!$ discrete syntactic monoids of a k element subset of A^* that are isomorphic to $S(W_1)$. By defining an equivalence relation θ on the set of all k element subsets of A^* by $(V_1, V_2) \in \theta$ if and only if $S(V_1) \cong S(V_2)$ we have $|W_{(1,n,k)}/\theta| \geq N_{(1,n,k)}/(|A|!)$. Let P be the property that a k element set of words W does not contain any unique maximally occurring letter a. Now $N_{(P,n,k)} \geq |W_{(P,n,k)}/\theta|$ so therefore

$$\frac{|W_{(P,n,k)}/\theta|}{|W_{(1,n,k)}/\theta|} < \frac{N_{(P,n,k)}}{(N_{(1,n,k)}/(|A|!))} = |A|!\frac{N_{(P,n,k)}}{N_{(1,n,k)}}$$

which tends toward 0 as n tends toward infinity. This combined with the above results shows that for any fixed positive integer k, almost all discrete syntactic monoids of k element languages in a finite alphabet are NFB. □

We note as a comparison that the results of [41] and [43] show that almost all semigroups (monoids) are in fact 3-nilpotent (3-nilpotent monoids) in the sense that the ratio of the number of 3-nilpotent semigroup operations (monoid operations) definable on an n element set to the number of all semigroup operations (monoid operations) definable on an n element set tends to 1 as n tends to infinity. It is easily shown that a 3-nilpotent semigroup or monoid must satisfy $xyx \approx xxy$ and therefore is FB by results from [70]. In fact a 3-nilpotent semigroup satisfies $x_1x_2x_3 \approx y_1y_2y_3$ and so generates a HFB variety with only finitely many subvarieties.

2.6 Joins of varieties generated by discrete syntactic monoids

Examples found by M. Volkov (see [82] for example) and M. Sapir [75] show that the class of finite FB semigroups and the class of finite NFB semigroups are not closed under taking direct products (or indeed of subsemigroups and homomorphic images). The properties of these examples appear to depend on the existence of

nontrivial subgroups. In this section we will address the problem of finding FB finite *aperiodic* semigroups whose direct product is NFB and NFB finite *aperiodic* semigroups whose direct product is FB. Note that Corollaries 2.2.6 and 2.4.12 above show that the class of finite FB aperiodic semigroups and the classes of finite FB and finite NFB aperiodic semigroups (and in particular the classes of FB or NFB discrete syntactic monoids of finite languages) are also not closed under taking subsemigroups or homomorphic images.

The following simple lemma is useful.

LEMMA 2.6.1 *[34] Let W_1 and W_2 be two sets of words over some alphabet X. Then $S(W_1 \cup W_2)$ satisfies the same identities as the direct product $S(W_1) \times S(W_2)$.*

DEFINITION 2.6.2 *For each $n > 1$ let $\mathsf{A_n}$ be the set of all words starting with a in the alphabet $\{ab, ba\}$ whose length is n (as words in this alphabet) and let A be a fixed element of $\mathsf{A_n}$, say $(ab)^{m_1}(ba)^{m_2}...(ab)^{m_k}$, where $m_i > 0$ for all $i < k$, $m_k \geq 0$, and $\Sigma_{i=1}^{k} m_i = n$.*

For $n > 2$, at least one of the words $(ab)^{n-1}ba$ and $ab(ba)^{n-1}$ is contained in the set $\mathsf{A_n}\backslash\{A\}$. Fix one of them that is contained in $\mathsf{A_n}\backslash\{A\}$ and call it B. For each $m \geq 1$ let ξ_m be a substitution defined by $\xi_m(ab) \equiv [Xm]$, $\xi_m(ba) \equiv [mX]$. We now construct an identity $L_{A,m} \approx R_{A,m}$ as follows. To make the word $L_{A,m}$, first replace every occurrence of ab in the word A by the word abt (where t is, as usual, a linear letter) and every occurrence of the word ba by the word bat. Let the resulting word be denoted by A'. Now replace every occurrence of a in A' by the letter x and every occurrence of b by corresponding occurrences of $\xi_m(ab)$ or $\xi_m(ba)$ from the word $\xi_m(B)$. That is, if the i^{th} letter to appear in B as a word in the alphabet $\{ab, ba\}$ is ab then the i^{th} occurrence of b in A' is to be replaced by $\xi_m(ab)$. Otherwise the i^{th} occurrence of b in A' is to be replaced by $\xi_m(ba)$. The same procedure is followed to make the word $R_{A,m}$ except each occurrence of b in A' is replaced with x and each occurrence of a is replaced with the corresponding subwords of $\xi_m(B)$. For example

let $n = 3$ and $A \equiv ababba$. So the only choice of B is the word $abbaba$. Now in this case A' is the word $abt_1abt_2bat_3$ and

$$L_{A,m} \equiv (x(x_1x_2 \ldots x_m)t_1)(x(x_m \ldots x_2x_1)t_2)((x_m \ldots x_2x_1)xt_3)$$

Likewise,

$$R_{A,m} \equiv ((x_1x_2 \ldots x_m)xt_1)((x_m \ldots x_2x_1)xt_2)(x(x_m \ldots x_2x_1)t_3).$$

LEMMA 2.6.3 *If* $\mathbf{S}$ *is a monoid for which the elements of* $\mathsf{A_n}\backslash\{A\}$ *(for some* $n > 2$*) are isoterms and for every* $m > 0$, $\mathbf{S} \models L_{A,m} \approx R_{A,m}$, *then* $\mathbf{S}$ *is NFB.*

Proof. If A is not the word $(ab)^n$ then by assigning a to x and b to respective linear letters t we find that $L_{A,m}(x, \tau)$ becomes the word $(ab)^n$ (recall that τ is the set of linear letters in a word). Since this is an isoterm for $\mathbf{S}$, $L_{A,m}(x, \tau)$ must be too. If $A \equiv (ab)^n$ then both $(ab)^{n-1}ba$ and $(ab)^{n-2}baab$ must be isoterms. By assigning a to x and maximal subwords of the form b^i to corresponding linear letters t we find that $xt_1xt_2...xt_{n-1}x$ and $xt_1xt_2...xt_{n-2}xxt_n$ are isoterms. These two facts combined ensure that $xt_1xt_2...xt_n$ is an isoterm. So for every non-linear letter y in $L_{A,m} \approx R_{A,m}$, the identity $L_{A,m}(y, \tau) \approx R_{A,m}(y, \tau)$ is a tautology and the words in this identity are isoterms for $\mathbf{S}$. Since B is an isoterm for $\mathbf{S}$, $L_{A,m}(x_1, x_2, ..., x_m)$ is an isoterm and for any $i \leq m$, $L_{A,m}(x, x_i)$ is essentially the word A (up to a change in letter names).

Let $L_{A,m} \approx w$ be any nontrivial identity satisfied by $\mathbf{S}$. The word $xt_1xt_2 \ldots xt_n$ is an isoterm for $\mathbf{S}$ so w differs from $L_{A,m}$ only by permutations within blocks. This means that for all $i \leq m$, $w(x, x_i)$ is equivalent up to a change in letter names to a word in $\mathsf{A_n}$ and for some $i \leq m$ the pair (x, x_i) is unstable in $L_{A,m} \approx w$ (the pair (x_i, x_j) must be stable in this identity since $L_{A,m}(x_i, x_j)$ is essentially the word B, an isoterm for $\mathbf{S}$). In fact since all words except A in $\mathsf{A_n}$ are isoterms for $\mathbf{S}$, $w(x, x_i)$ must be equivalent up to change in letter names to the word A and so for every

$k \leq n$, the pair $({}_k x, {}_k x_i)$ is unstable in $L_{A,m} \approx w$. Because $w(x, x_i) \not\equiv L_{A,m}(x, x_i)$ it must be the case that $w(x, x_i) \equiv R_{A,m}(x, x_i)$. We now show that $w \equiv R_{A,m}$.

Without loss of generality we may assume that B is the word $(ab)^{n-1}ba$. The first letter of A is the letter a so it follows that $xx_1x_2 \ldots x_i \ldots x_m$ is an initial segment of $L_{A,m}$. Since $w(x, x_i) \equiv R_{x,x_i}$ and $L_{A,m}(x_1, x_2, \ldots x_m) \equiv w(x_1, x_2, \ldots x_m)$ is an isoterm, the word $x_1x_2 \ldots x_i x$ is an initial segment of w. This means that (x, x_1) is an unstable pair in $L_{A,m} \approx w$. Indeed, as discussed above, this implies that for every $k \leq n$, the pair $({}_k x, {}_k x_1)$ is unstable in $L_{A,m}$. Now because $A \not\equiv B$, one of the words $xx_m \ldots x_1$ or $x_1 \ldots x_m x$ is a subword of $L_{A,m}$. However (x, x_1) is unstable in $L_{A,m} \approx w$ and $L_{A,m}(x_1, \ldots x_m)$ is an isoterm for $\mathbf{S}$ so (x, x_n) is also an unstable pair. Again this means that for every $k \leq n$, $({}_k x, {}_k x_n)$ is unstable. It is now evident from the fact that $L_{A,m}(x_1, \ldots, x_m)$ is an isoterm for $\mathbf{S}$ that $w \equiv R_{A,m}$.

Now we show that there is no derivation of $L_{A,m} \approx R_{A,m}$ involving identities of $\mathbf{S}$ that contain less than n letters. Assume otherwise. There is an identity $p \approx q$ involving fewer than n letters and a substitution θ such that $L_{A,m} \equiv u\theta(p)v$ and $R_{A,m} \equiv u\theta(q)v$. By the choice of B we can assume without loss of generality that there is only one occurrence of the subword $x_{i+1}x_i$ in $L_{A,m}$, say the j^{th} occurrence. Since we are assuming that $|c(p)| < n$ there must be a linear letter z in $c(p)$ such that $\theta(z)$ contains $x_{i+1}x_i$ as a subword. There is also a letter $x' \in c(p)$ whose k^{th} occurrence (for some k) is assigned by θ the j^{th} occurrence of x in $L_{A,m}$. By the structure of $R_{A,m}$ it follows that $({}_k x', z)$ is unstable in $p \approx q$ and $p \approx q$ can be deleted to the identity

$$x't_1 \ldots ({}_{j'}x')z(t_{j'})x't_{j'+1} \ldots x't_{(occ(x',p))} \approx x't_1 \ldots z({}_{j'}x')(t_{j'})x't_{j'+1} \ldots x't_{(occ(x',p))}.$$

Since $xt_1 \ldots xt_n$ is an isoterm and $occ(x', p) \leq n$, by Lemma 2.3.1 the left hand side of this is an isoterm, a contradiction. Thus no such identity $p \approx q$ exists. Therefore any basis for $\mathbf{S}$ must contain identities involving arbitrarily large numbers of letters and is therefore infinite. □

Recall that W_n is the set of all words in the alphabet $\{a, b\}$ with at most n

occurrences of any letter. For any fixed word A from A_n with $n > 1$ let $W_{A,n}$ be the result of removing from W_n the word A and the word $\bar{A}$ obtained from A by simultaneously replacing a by b and b by a.

COROLLARY 2.6.4 *For $n > 1$, $S(W_{n,A})$ is NFB.*

Proof: For $n > 2$ Lemma 2.6.3 can be used as follows. Since $\mathsf{A}_n \backslash \{A, \bar{A}\}$ is a subset of $W_{A,n}$, every word in $\mathsf{A}_n \backslash \{A, \bar{A}\}$ is an isoterm for $S(W_{A,n})$. On the other hand, every word in $W_{A,n}$ has length less than $2n + 1$. So if θ is a substitution such that $\theta(L_{A,m})$ is contained in $W_{A,n}$ then $\theta(L_{A,m})$ must have length less than $2n + 1$. Therefore θ must assign 1 to all but at most two letters from $\{x\} \cup \{x_i; i \leq m\}$. In this case either $\theta(L_{A,m}) \equiv \theta(R_{m,A})$ or $\theta(L_{A,m})$ is equivalent up to a change of letter names to A and $\theta(R_{A,m})$ is similarly equivalent to $\bar{A}$. Since $S(W_{A,n}) \models A \approx \bar{A}$, $S(W_{A,n}) \models L_{A,m} \approx R_{A,m}$ for every $m > 1$. Therefore by Lemma 2.6.3, $S(W_{A,n})$ is NFB.

For $n = 2$, A_n is the set $\{abab, abba\}$. In this case $S(W_n)$ is equationally equivalent to $S(\{abab, abba, aabb\})$ since $\{abab, abba, aabb\}$ contains a copy (up to a change of letter names) of every 2-limited word in a two letter alphabet. Thus to prove the result we need to show that $S(\{abab, aabb\})$ and $S(\{abba, aabb\})$ are NFB. For the first of these cases we can apply Lemma 2.3.3. The second case is due to O. Sapir and follows from a similar lemma in [34] or [79]. For example $xytxy$ is an isoterm for $S(\{abab, aabb\})$ since $xyxy$ and xyx are. However for any unstable pair of letters (x, y) in the identity $L_n \equiv [X2n]t[\mathcal{X}2n] \approx [\mathcal{X}2n]t[X2n] \equiv R_n$, the identity $L_n(x, y) \approx R_n(x, y)$ is the identity $xyyx \approx yxxy$ which is a satisfied by $S(\{abab, aabb\})$. Thus $S(\{abab, aabb\}) \models L_n \approx R_n$ for every $n > 0$ and by Lemma 2.3.3, is NFB. The Corollary is proved. □

The description in [80] of all words w in a two letter alphabet $\{a, b\}$ for which $S(\{w\})$ is NFB (see Theorem 2.0.10 of this thesis) shows that for any word A chosen from A_n, the syntactic monoid $S(\{A\})$ is NFB. The following corollary now follows Corollary 2.2.2 and Corollary 2.6.4.

COROLLARY 2.6.5 *For every $n > 1$ and every word $A \in \mathsf{A}_n$, the monoids $S(\{A\})$ and $S(W_{n,A})$ are NFB but $S(\{A\}) \times S(W_{n,A})$ and $S(\{A\} \cup W_{n,A})$ are FB.*

Since $S(\{abab\})$ and $S(\{abba\})$ are NFB, one might wonder if $S(\{abab, abba\})$ is FB, therefore giving a smaller example. Example 2.3.8 shows however that this is not true. Nevertheless, we can find two words w_1 and w_2 such that $S(\{w_1\})$ and $S(\{w_2\})$ are NFB but $S(\{w_1, w_2\})$ is FB. First consider the following lemma.

LEMMA 2.6.6 *If w is an isoterm for a monoid $\mathbf{S}$ then $Id(\mathbf{S}) \subseteq Id(S(\{w\}))$.*

Proof: Let $p \approx q$ be an identity not satisfied by $S(\{w\})$. This means that there is a substitution θ such that $\theta(p)$ is a subword of w and $\theta(p) \not\equiv \theta(q)$. So $w \equiv u\theta(p)v$ for some words u and v so that $u\theta(p)v \not\equiv u\theta(q)v$. But then $p \approx q \vdash u\theta(p)v \approx u\theta(q)v$ so w is not an isoterm for any semigroup satisfying $p \approx q$. That is, $\mathbf{S} \not\models p \approx q$. The lemma is proved. □

Let w be the word *ababcddee*. Since $S(\{w\})$ contains the subsemigroup $S(\{abab\})$ and the subsemigroup $S(\{ddee\})$, $Id(S(\{w\}))$ is contained in both $Id(S(\{abab\}))$ and $Id(S(\{aabb\}))$ and therefore also in $Id(S(\{abab, aabb\}))$. On the other hand since $xyxy$, $xyyx$, xxy, yxx and xyx are all isoterms for $S(\{abab, aabb\})$, so must be the word w and therefore Lemma 2.6.6 shows that

$$Id(S(\{ababcddee\})) \supseteq Id(S(\{abab, aabb\})).$$

We can conclude that the monoid $S(\{ababcddee\})$ is equationally equivalent to the monoid $S(\{abab, aabb\})$. In a similar way one can show that $S(\{ababcddee, abba\})$ is equationally equivalent to $S(\{abab, aabb, abba\})$. Combining these ideas we obtain the following example.

EXAMPLE 2.6.7 *The monoids $S(\{ababcddee\})$ and $S(\{abba\})$ are NFB but the monoid $S(\{ababcddee, abba\})$ is FB*

Another simple example is the following.

EXAMPLE 2.6.8 *The monoids $S(\{abcba\})$ and $S(\{abcab\})$ are NFB but the monoid $S(\{abcba, abcab\})$ is FB.*

Proof: The argument used in Example 2.3.4 applies equally well to the monoid $S(\{abcab\})$ and likewise a similar lemma from [34] due to O. Sapir may be used in the case of $S(\{abcba\})$. So $S(\{abcab\})$ and $S(\{abcab\})$ are NFB. On the other hand in Theorem 2.2.11 above it was shown that $S(\{abcab, abcba\})$ (and by Lemma 2.6.6, $S(\{abcabdefgfe\})$) is FB. □

The relevance of this example is due to the following theorem.

THEOREM 2.6.9 *For any $n \geq 2$ the monoids $S(\{abcab, abcba\})$ and $S(\{a^n b^n\})$ are FB but the monoids $S(\{abcab, abcba\}) \times S(\{a^n b^n\})$ and $S(\{abcab, abcba, a^n b^n\})$ are NFB.*

Proof: Theorem 2.2.11 shows that $S(\{abcab, abcba\})$ is FB and $S(\{a^n b^n\})$ is FB by Theorem 2.0.10. Example 2.3.6 shows that $S\{abcab, abcba, a^n b^n\})$ is NFB. □

Thus by Lemma 2.6.1 and this theorem we have an example of two finite FB aperiodic semigroups whose direct product is NFB. The problem of finding such an example was raised by M. Sapir about 10 years ago.

Chapter 3

Small INFB finite semigroups.

As discussed in the historical overview, a powerful algorithmic description of the class of finite INFB semigroups has been obtained by M. Sapir [73], [74]; see Theorem 1.1.1 and Theorem 1.1.2 of this thesis.

The power of these theorems is demonstrated by the following simple example [73]. Consider the monoid $\mathbf{B}_2^1$ with semigroup presentation $\langle 1, a, b : a^2 = b^2 = 0, aba = a, bab = b\rangle$; clearly $\mathbf{B}_2^1$ has period 1, ab is idempotent and a divides ab. However both $ab(a)ab$ and $ab(a^2)ab$ equal 0 in $\mathbf{B}_2^1$ and 0 is not an element of the maximal subgroup containing ab. Therefore by Theorem 1.1.2, $\mathbf{B}_2^1$ is INFB.

In what follows it will frequently be necessary to consider pairs of the form (a, e) where a and e are elements of a semigroup $\mathbf{S}$, e is idempotent and a divides e. Such a pair will be called a *dividing pair* and we will say *INFB occurs at* (a, e) if this pair satisfies the conditions of part (ii) of Theorem 1.1.2 for some submonoid of $\mathbf{S}$ containing a and e. Recall also that $\mathbf{S}_e$ is the maximal subgroup of $\mathbf{S}$ containing e (see Theorem 1.1.2).

The semigroup $\mathbf{B}_2^1$ is a particularly important example of a finite INFB semigroup since it generates a variety that is minimal amongst those generated by finite INFB semigroups [74]. In particular if $\mathbf{S}$ is any semigroup that has only nilpotent subgroups (such as an aperiodic semigroup) then $\mathbf{S}$ is INFB if and only if $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$,

the variety generated by $\mathbf{S}$ [74].

DEFINITION 3.0.10 *Let* $\mathbf{M}$ *be a finite INFB semigroup from a variety* $\mathcal{V}$. *If there is a subvariety* $\mathcal{V}'$ *of* $\mathcal{V}$ *containing* $\mathbf{M}$ *so that for every semigroup* $\mathbf{S} \in \mathcal{V}'$, $\mathbf{S}$ *is INFB if and only if* $\mathbf{M} \in \mathcal{V}(\mathbf{S})$ *then* $\mathbf{M}$ *will be said to be a minimal finite INFB semigroup for* $\mathcal{V}$ *and the minimum finite INFB semigroup for* $\mathcal{V}'$. *The variety* $\mathcal{V}(\mathbf{M})$ *will be called a minimal finitely generated INFB variety. Similarly if* C *is a class of semigroups (not necessarily a variety) containing* $\mathbf{M}$ *and for every* $\mathbf{S} \in C$, $\mathbf{S}$ *is INFB if and only if* $\mathbf{M} \in \mathcal{V}(\mathbf{S})$ *then* $\mathbf{M}$ *will also be called the minimum INFB semigroup for* C.

So in the terminology of this definition, $\mathbf{B}_2^1$ generates a minimal finitely generated INFB variety and generates a variety that is the minimum INFB variety for the class of finite semigroups with only nilpotent subgroups. In this chapter we use the theorems of [73] and [74] to find some other classes for which $\mathcal{V}(\mathbf{B}_2^1)$ is the minimum INFB variety and, modulo certain properties of completely simple semigroups, we give a description of all minimal finite INFB divisors. In connection with the results of the previous chapter, it is interesting to note that $\mathbf{B}_2^1$ is in fact the syntactic monoid of the language $\{ab\}^*$.

3.1 Classes for which $\mathcal{V}(\mathbf{B}_2^1)$ is the minimum INFB variety

We first recall an extract of a result that is central to the arguments used in [74] (proved partly by M. Sapir in [74] and partly by L. Shevrin [83], [84] and [85]).

LEMMA 3.1.1 *Let* $\mathbf{S}$ *be a finite monoid. If there is no homomorphic image of a submonoid of* $\mathbf{S}$ *isomorphic to* $\mathbf{B}_2^1$ *or* $\mathbf{A}_2^1$ *then for every idempotent* $e \in \mathbf{S}$ *and every element* a *dividing* e *in* $\mathbf{S}$ *the element* eae *belongs to* $\mathbf{S}_e$. *Furthermore if for every idempotent* $e \in \mathbf{S}$ *and every element* a *dividing* e *in* $\mathbf{S}$ *the element* eae *belongs to* $\mathbf{S}_e$

then for every idempotent $f \in \mathbf{S}$ and any element b dividing f in $\mathbf{S}$, the element b^2 divides f in $\mathbf{S}$.

We now have enough information to prove the following simple theorem, effectively a corollary of Theorems 1.1.1 and 1.1.2.

THEOREM 3.1.2 *If $\mathbf{S}$ is a finite regular semigroup with period d then the following are equivalent:*
(i) $\mathbf{S}$ is INFB,
(ii) $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$,
(iii) $S(\{a\}) \in \mathcal{V}(\mathbf{S})$, where $S(\{a\})$ is the three element monoid with presentation $\langle 1, a; aa = 0\rangle$,
(iv) $\mathbf{S} \not\models xyx \approx (xy)^{d+1}x$.

Proof: The implications (ii)⇒(i) and (ii)⇒(iii)⇒(iv) follow immediately since $\mathbf{B}_2^1$ is INFB, $S(\{a\}) \in \mathcal{V}(\mathbf{B}_2^1)$ and $S(\{a\}) \not\models xyx \approx (xy)^{d+1}x$ for any $d > 0$. Implication (i)⇒(iv) follows since if $\mathbf{S} \models xyx \approx (xy)^{d+1}x$ the Zimin word Z_2 is not an isoterm for $\mathbf{S}$ and by Theorem 1.1.1, $\mathbf{S}$ is not INFB.

We now show that condition (iv) implies condition (ii). Say that the identity $xyx \approx (xy)^{d+1}x$ fails on the finite regular semigroup $\mathbf{S}$. So there are elements a and b of $\mathbf{S}$ for which $aba \neq (ab)^{d+1}a$. Since $\mathbf{S}$ is regular there is an idempotent e with $e\mathcal{R}a$ and $ea = a$. So $aba = (eabe)a$ and $(ab)^{d+1}a = (eab)^{d+1}ea = (eabe)^{d+1}a$. Now consider the monoid $e\mathbf{S}e$. This is a regular monoid since for any element $exe \in e\mathbf{S}e$ with inverse x' in $\mathbf{S}$, $exe = (exe)x'(exe) = (exe)(ex'e)(exe)$. If $e\mathbf{S}e$ is completely regular then it satisfies $x \approx x^{d+1}$. In this case $eabe = (eabe)^{d+1}$ and therefore $aba = eabea = (eabe)^{d+1}a = (ab)^{d+1}a$, a contradiction. Therefore $e\mathbf{S}e$ is not completely regular and there is an element $c \in e\mathbf{S}e$ which does not lie in a subgroup of $e\mathbf{S}e$. Consider the $\mathcal{D}$-class D_c of c in $e\mathbf{S}e$. The principle factor $\mathbf{P}$ of D_c is a completely 0-simple semigroup in which $c^2 = 0$. Since D_c is regular there is a (non zero) idempotent $f \in D_c$ so that c divides f. However in $\mathbf{P}$ we have $c^2 = 0$ and so

c^2 does not divide f in $\mathbf{P}$ or therefore in $e\mathbf{S}e$ and so by Lemma 3.1.1, at least one of the INFB monoids $\mathbf{B}_2^1$ or $\mathbf{A}_2^1$ is contained in $\mathcal{V}(\mathbf{S})$ and $\mathbf{S}$ is INFB. The result now follows since $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{A}_2^1)$ (see page 6). □

Theorem 3.1.2 is particularly surprising when one considers the existence of finite INFB semigroups not generating varieties containing $\mathbf{B}_2^1$ and the fact that every (finite) semigroup is embeddable in a (finite) regular semigroup (of course this embedding involves a comparatively large semigroup of all transformations of a set). The situation is emphasised by the following corollary of Theorem 3.1.2.

COROLLARY 3.1.3 *A finite monoid* $\mathbf{S}$ *is embeddable in a finitely based finite regular semigroup only if* $\mathbf{S}$ *is regular.*

Proof: The statement follows because a finite monoid containing a non group element generates a variety containing $S(\{a\})$. □

It would be interesting if the reverse implication also held true for WFB monoids and to obtain a corresponding theorem for finite semigroups without an identity element. Since any semigroup satisfying $xyx \approx (xy)^{d+1}x$ satisfies $x^3 \approx x^{2d+3}$, Theorem 3.1.2 implies that no semigroup with index greater than three can be embedded in a finitely based finite regular semigroup. In fact we can reduce these bounds further.

PROPOSITION 3.1.4 *If* $\mathbf{S}$ *is a semigroup with index greater than two, then* $\mathbf{S}$ *is not embeddable into a finite finitely based regular semigroup.*

Proof: Assume that $\mathbf{S}$ is embedded in a finite regular semigroup $\mathbf{R}$. There is an element $a \in \mathbf{S} \subseteq \mathbf{R}$ so that $a^2 \neq a^{2+i}$ for any $i > 0$. Since $\mathbf{R}$ is regular there is an idempotent e so that $ea = a$. The element eae cannot lie in a subgroup of $\mathbf{R}$ since then for some d we have $a^2 = (ea)(ea) = (eae)a = (eae)^{d+1}a = a^{d+2}$. Therefore the monoid $e\mathbf{R}e$ is INFB since $S(\{a\}) \in \mathcal{V}(e\mathbf{R}e)$. □

Note that there are many WFB and even FB regular semigroups with index equal to two (an example is $\mathbf{B}_2 = \mathbf{B}_2^1 \backslash \{1\}$).

Recall that an orthodox semigroup is a regular semigroup whose idempotents form a subsemigroup. Orthodox semigroups are a well known and important generalisation of inverse semigroups. Rasin [71] has showed that a finite orthodox completely regular semigroup is HFB and Question 8.2 of [82] asks whether a finite orthodox semigroup $\mathbf{S}$ is finitely based if and only if $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$. Combining the result of Rasin with Theorem 3.1.2 we get the following partial solutions to this question.

COROLLARY 3.1.5 *A finite orthodox monoid is FB if and only if it is HFB and if and only if it is not INFB. A finite orthodox semigroup* $\mathbf{S}$ *is INFB if and only if* $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$.

In the class of monoids therefore, Question 8.2 of [82] has a positive solution. Recalling the examples of Chapter 2 (see Theorem 2.5.11 for example) we see that there are a large number of WNFB finite semigroups whose idempotents form a subsemigroup, even a subsemilattice but are not regular. Therefore if the condition of regularity is removed from the definition of an orthodox semigroup the first sentence of Corollary 3.1.5 no longer holds. The second sentence however does continue to hold.

THEOREM 3.1.6 *If the idempotents of a finite semigroup* $\mathbf{S}$ *form a subsemigroup of* $\mathbf{S}$ *then* $\mathbf{S}$ *is INFB if and only if* $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$.

Proof: If the idempotents of a semigroup $\mathbf{S}$ form a subsemigroup then for every idempotent e, the idempotents of the submonoid $e\mathbf{S}e$ also form a subsemigroup of $e\mathbf{S}e$. Therefore by Theorem 1.1.2 we need only consider the case when $\mathbf{S}$ is a monoid.

If $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$ then $\mathbf{S}$ is INFB by the definition of being inherently nonfinitely based. Assume that $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$. Since $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{A}_2^1)$, by Lemma 3.1.1 for every dividing pair (a, e), $eae \in \mathbf{S}_e$. Now for any $i > 0$, a^i divides a^{2^i} and by Lemma 3.1.1, a^{2^i} divides e. Therefore $ea^ie \in \mathbf{S}_e$ for all $i > 0$. We now use induction to show

that $ea^ie = (eae)^i$. From this it follows that $ea^{d+1}e = (eae)^{d+1} = eae \in \mathbf{S}_e$, and by Theorem 1.1.2, $\mathbf{S}$ is not INFB.

For any $g \in \mathbf{S}_e$, let g^{-1} denote the group inverse of g in $\mathbf{S}_e$. Now for any $k > 0$ we have that $(eae)^{-1}a$ and $a(eae)^{-1}$ are both idempotent since, for example,

$$(eae)^{-1}a(eae)^{-1}a = (eae)^{-1}eae(eae)^{-1}a = (eae)^{-1}a.$$

Therefore $(eae)^{-1}aa(eae)^{-1} = (eae)^{-1}ea^2e(eae)^{-1}$ is idempotent and since

$$(eae)^{-1}ea^2e(eae)^{-1} \in \mathbf{S}_e,$$

$(eae)^{-1}ea^2e(eae)^{-1} = e$. Therefore $ea^2e = (eae)^2$.

Now assume that $ea^ke = (eae)^k$. Since $(eae)^{-1}a$ and $a^k(ea^ke)^{-1}$ are idempotent, so is the element $(eae)^{-1}aa^k(ea^ke)^{-1}$. Therefore

$$(eae)^{-1}aa^k(ea^ke)^{-1} = (eae)^{-1}ea^{k+1}e(ea^ke)^{-1} = (eae)^{-1}ea^{k+1}e(eae)^{-k} = e.$$

Therefore $ea^{k+1}e = (eae)^{k+1}$ as required. In particular $ea^{d+1}e = (eae)^{d+1} = eae$ since the exponent of $\mathbf{S}_e$ divides the period, d, of $\mathbf{S}$. □

By a well known result from [4] the class of all finite semigroups whose idempotents form a subsemigroup is exactly the psuedovariety generated by the class of finite orthodox semigroups.

Theorem 1.1.2 also provides a way of increasing the power of this result.

DEFINITION 3.1.7 *If P is a property of semigroups then a semigroup $\mathbf{S}$ has the property P locally or $\mathbf{S}$ is locally-P, if for every idempotent e, $e\mathbf{S}e$ has the property P.*

COROLLARY 3.1.8 *If P is a property so that the finite semigroups with P are INFB if and only if $\mathbf{B}_2^1$ is contained in the variety they generate then a finite locally-P semigroup is INFB if and only if $\mathbf{B}_2^1$ is contained in the variety it generates.*

Proof: Let $\mathbf{S}$ be a finite locally-P semigroup. Then by Theorem 1.1.2, $\mathbf{S}$ is INFB if and only if $e\mathbf{S}e$ is INFB for some idempotent $e \in \mathbf{S}$. The semigroup $e\mathbf{S}e$ has the property P and therefore is INFB if and only if $\mathbf{B}_2^1 \in \mathcal{V}(e\mathbf{S}e)$. Since $e\mathbf{S}e$ is a subsemigroup of $\mathbf{S}$ the result follows. □

3.2 Number of elements in a minimal finite INFB semigroup

In this section we address the possible size of an INFB semigroup $\mathbf{S}$ for which $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$. We will assume throughout that $\mathbf{S}$ is a finite INFB monoid of period d with $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$ and that INFB occurs at the dividing pair (a, e). As in the previous section $\mathbf{S}_e$ will denote the largest subgroup of $\mathbf{S}$ containing e and $ea^ie \in \mathbf{S}_e$ for every $i \geq 0$. A number of simple lemmas will lead to a lower bound for the cardinality of an INFB semigroup $\mathbf{S}$ with $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$.

LEMMA 3.2.1 *No subgroup of* $\mathbf{S}$ *contains* a.

Proof: If a were in a subgroup of $\mathbf{S}$ then $a = a^{d+1}$ and $eae = ea^{d+1}e$ contradicting the fact that INFB occurs at (a, e). □

LEMMA 3.2.2 *Let* s *and* t *be elements of* $\mathbf{S}_e$ *and* $i \geq 0$.
(i) The elements as, sa *are not contained in* $\mathbf{S}_e$,
(ii) $sa^i = ta^i \Rightarrow s = t$,
(iii) $sa^i \neq at$ *and* $sa \neq a^it$.

Proof: (i) If $sa \in \mathbf{S}_e$ then $rs^{-1}sa = ra \in \mathbf{S}_e$ for any $r \in \mathbf{S}_e$. Say $ea = r$ for some $r \in \mathbf{S}_e$. Then $ea^{d+1}e = rea^de = r^2ea^{d-1}e = \ldots = r^deae = eae$, contradicting the fact that INFB occurs at (a, e). That $as \notin \mathbf{S}_e$ follows by symmetry.

(ii) Say $sa^i = ta^i$. Then

$$\begin{aligned} s &= s(ea^ie)(ea^ie)^{-1} \\ &= (sa^i)e(ea^ie)^{-1} \\ &= (ta^i)e(ea^ie)^{-1} \\ &= t(ea^ie)(ea^ie)^{-1} \\ &= t. \end{aligned}$$

(iii) Say $sa^i = at$ for some s, $t \in \mathbf{S}_e$. So $eae = eatt^{-1} = sa^it^{-1} = att^{-1} = ae$. But $eae \in \mathbf{S}_e$ and $ae \notin \mathbf{S}_e$ by part (i), a contradiction. The case $sa = a^it$ follows by symmetry. □

LEMMA 3.2.3 *Let s and t be arbitrary elements of $\mathbf{S}_e$. Then $sa^2 \neq ta$ and $a^2s \neq at$.*

Proof: Say $sa^2 = ta$. Then $sa^{d+1} = sa^2a^{d-1} = taa^{d-1} = ts^{-1}sa^d = ts^{-1}(sa^2)a^{d-2} = ts^{-1}ta^{d-1} = \ldots = (ts^{-1})^{d-1}ta = (ts^{-1})^{-1}ta = st^{-1}ta = sa$. Therefore $eae = s^{-1}sae = s^{-1}sa^{d+1}e = ea^{d+1}e$, contradicting the fact that INFB occurs at (a, e). That $a^2s \neq at$ follows by symmetry. □

LEMMA 3.2.4 *For every $s \in \mathbf{S}_e$, we have $sa^2 \notin \mathbf{S}_e$ and $a^2s \notin \mathbf{S}_e$.*

Proof: Assume $sa^2 \in \mathbf{S}_e$. So $s^{-1}sa^2 = ea^2 \in \mathbf{S}_e$ and therefore $ea^2 = ea^2e$. Let i and p be the index and period respectively of the subsemigroup $\langle a \rangle$ of S generated by a.

Case 1. p is odd.

If i is odd then $a^{i+1} = a^{i+1+p}$ and $i + 1$ is even. Let $2j$ be the even element of $\{i, i + 1\}$ (that is, j is the integer part of $(i + 1)/2$). So $ea^{2j} = ea^2a^{2j-2} = ea^2ea^{2j-2} = \ldots = (ea^2)^j \in \mathbf{S}_e$. But since p is odd, $ea^{2j} = ea^{2j+p} = ea^{2j}(a^2)^{(p-1)/2}a = (ea^2)^j(a^2)(a^2)^{(p-1)/2-1}a = \ldots = (ea^2)^{j+(p-1)/2}a$. Now $ea^2 \in \mathbf{S}_e$ and by Lemma 3.2.2,

$sa \notin \mathbf{S}_e$, so therefore $(ea^2)^{j+p/2-1/2}a \notin \mathbf{S}_e$, contradicting the fact that $(ea^2)^{j+(p-1)/2}a = ea^{2j} = (ea^2)^j \in \mathbf{S}_e$.

Case 2. p is even.

Since p divides d and p is even, $\frac{p}{2}$ divides $\frac{d}{2}$ and d is even. Therefore if for some $s \in \mathbf{S}_e$, we have $s^{p/2} = e$ then $s^{d/2} = e$. Now since $ea^2 \in \mathbf{S}_e$,

$$ea^{d+1}e = ea^2a^{d-1}e = ea^2eea^{d-1}e = \ldots = (ea^2)^{d/2}eae.$$

We now show that $(ea^2)^{p/2} = e$ and therefore $ea^{d+1}e = eae$, a contradiction as required.

Let $2j$ be the even element of $\{i, i+1\}$. So $ea^{2j} = ea^2a^{2j-2} = \ldots = (ea^2)^j \in \mathbf{S}_e$. But

$$ea^{2j} = ea^{2j+p} = ea^2(a^2)^{j+p/2-1} = ea^2e(a^2)^{j+p/2-1} = \ldots = (ea^2)^{j+p/2}$$

Therefore $(ea^2)^j = (ea^2)^{j+p/2} = (ea^2)^j(ea^2)^{p/2}$ and so $(ea^2)^{p/2} = e$ as required.

Therefore sa^2 is not contained in $\mathbf{S}_e$. That a^2s is not contained in $\mathbf{S}_e$ follows by symmetry. □

LEMMA 3.2.5 *For any elements $s, t \in \mathbf{S}_e$, $sa^2 \neq a^2t$.*

Proof: If $sa^2 = a^2t$ then $sa^2e = a^2te = a^2t$. But by Lemma 3.2.4, $a^2t \notin \mathbf{S}_e$, contradicting the fact that $sa^2e \in \mathbf{S}_e$. □

LEMMA 3.2.6 *If $i, j \in \{1, 2\}$ and $s \in \mathbf{S}_e$ then $a^isa^j \notin \mathbf{S}_e$.*

Proof: Say $a^isa^j \in \mathbf{S}_e$ and let $t = ea^is \in \mathbf{S}_e$. Then $a^isa^j = ea^isa^j = ta^j$, a contradiction since ta^j is not an element of $\mathbf{S}_e$ by Lemmas 3.2.2 part (i) and 3.2.4. □

LEMMA 3.2.7 *If $i, j, k, l \in \{1, 2\}$ and $s, t \in \mathbf{S}_e$ then $a^isa^j = a^kta^l$ implies $s = t$, $i = k$, $j = l$.*

Proof: Say $j \neq l$. Without loss of generality we may assume $j = 1$ and $l = 2$. Then $a^i sa = a^k ta^2$ and so $(ea^i s)a = (ea^k t)a^2$, contradicting Lemma 3.2.3. Therefore, by symmetry, $i = k$ and $j = l$. So $a^i sa^j = a^i ta^j$ and therefore $ea^i sa^j e = ea^i ta^j e$. So $s = t$ as required. □

LEMMA 3.2.8 *If $i, j, k \in \{1, 2\}$ and $s, t \in \mathbf{S}_e$ then $a^i sa^j \neq a^k t$ or ta^k.*

Proof: If $a^i sa^j = a^k t$ then $ea^i sa^j = ea^k t$, contradicting Lemmas 3.2.2 (i) and 3.2.4. Likewise, by symmetry, $a^i sa^j \neq ta^k$. □

LEMMA 3.2.9 *For any $s \in \mathbf{S}_e$,*

$$a \notin \{s, sa, as, saa, aas, asa, aasa, asaa, aasaa, 1\}.$$

Proof: Firstly $a \neq 1$ since otherwise $eae = ea^{d+1}e \in \mathbf{S}_e$. Secondly for any $i, j \in \{0, 1, 2\}$, $(a^i sa^j)^{d+1} = a^i(sa^{i+j}e)^d sa^j = a^i sa^j$. Since $eae \neq ea^{d+1}e$, the result follows. □

LEMMA 3.2.10 *For any $i, j \in \{0, 1, 2\}$ and $s \in \mathbf{S}_e$, $1 \neq a^i sa^j$.*

Proof: If $i > 0$, $1 \neq a^i sa^j$ since then $e = e1 = a^i sa^j e$, contradicting Lemmas 3.2.2 (i) and 3.2.4. By symmetry the only remaining case is when $i = j = 0$, that is when $1 = s \in \mathbf{S}_e$. This is impossible since $a = a1 \neq as$ by Lemma 3.2.2 (i). □

Combining Lemmas 3.2.2 through 3.2.10 we have the following.

THEOREM 3.2.11 *The sets $\{1\}$, $\{a\}$, $\{a^i sa^j : s \in \mathbf{S}_e,\ i, j \leq 2\}$ are disjoint in $\mathbf{S}$.*

COROLLARY 3.2.12 *If $\mathbf{T}$ is a semigroup with $|T| < 56$ then $\mathbf{T}$ is INFB if and only $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{T})$.*

Proof: If $\mathbf{S}$ is a finite INFB semigroup and $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$ then $eae \in \mathbf{S}_e$ for every dividing pair (a, e). By Theorem 1.1.2, for one such dividing pair (a, e), eae and $ea^{d+1}e$ do not lie in the same coset of $\mathbf{S}_e$ modulo $\Gamma(\mathbf{S}_e)$ (recall that if $\mathbf{G}$ is a group then $\Gamma(\mathbf{G})$ is the upper hypercentre of $\mathbf{G}$). Since both these elements are contained in $\mathbf{S}_e$, we must have $\Gamma(\mathbf{S}_e) \neq \mathbf{S}_e$. If $\mathbf{G}$ is a group then by definition, $\Gamma(\mathbf{G}) = \mathbf{G}$ exactly when $\mathbf{G}$ is nilpotent. The smallest non nilpotent group $\mathbf{G}$ is the six element centreless group $\mathbf{S}_3$ with upper hypercentre equal to $\{1\}$. By Theorem 3.2.11 there is a disjoint copy of $\mathbf{S}_e$ for each pair $\{(i,j); i,j \in \{0,1,2\}\}$ (that is, nine copies of $\mathbf{S}_e$) as well as an element 1 and the element a. This sets the minimum size for such a semigroup as $9 \times 6 + 1 + 1 = 56$. □

As will be shown in the following section, there do exist quite a few INFB semigroups $\mathbf{S}$ with 56 elements and with $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$, so this bound is the best possible. A second corollary of Theorem 3.2.11 also follows.

COROLLARY 3.2.13 *If* $\mathbf{S}$ *is a semigroup with at most* 8 *non-nilpotent subgroups then* $\mathbf{S}$ *is INFB if and only if* $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$.

3.3 Minimal INFB divisors for finite semigroups

We now describe two constructions for making finite INFB monoids generating varieties not containing $\mathbf{B}_2^1$. These constructions will be based around finite centreless groups. The importance of centreless groups here lies in the fact that the upper hypercentre of a group $\mathbf{G}$ is a normal subgroup $\Gamma(\mathbf{G})$ such that $\mathbf{G}/\Gamma(\mathbf{G})$ is centreless.

Throughout the remainder of this chapter it will be convenient to consider (contrary to the usual convention) the ij^{th} entry of a matrix as the entry in the $(i+1)^{th}$ row and the $(j+1)^{th}$ column. For example the first entry in any matrix will be the 00^{th} entry and a Rees matrix semigroup (without 0 element) $\mathcal{M}(\mathbf{G}, m, n, P)$ over a group $\mathbf{G}$ with $n \times m$ matrix P will be considered as a set of the form

$$\{(i, g, j) : g \in \mathbf{G}, 0 \leq i \leq m-1, 0 \leq j \leq n-1\}$$

with multiplication $(i,g,j)(i',g',j') = (i, gP_{j,i'}g', j')$, where $P_{i,j}$ is the ij^{th} entry of the matrix P (according to the altered convention above). If a is a non group element of a monoid $\mathbf{S}$ we will let a^0 denote the identity element 1 of $\mathbf{S}$.

DEFINITION 3.3.1 *Let $\mathbf{G}$ be a finite centreless group with identity element e and exponent d. Let g and g_1 be (possibly identical) elements of the group $\mathbf{G}$. Construct a 3×3 matrix with group entries as follows: let $P_{2,2} = g$ and let g_i denote the element $(g_1g)^{i-1}g_1$; let h be any element of $\mathbf{G}\backslash\{g_2^{-1}g_1g_2^{-1}\}$; and for $i,j \leq 2$ define*

$$P_{i,j} = \begin{cases} e, \ if \ i = j = 0 \\ h, \ if \ i + j = 1 \\ g_2^{-1}g_{i+j}g_2^{-1}, \ if \ i + j \geq 2. \end{cases}$$

Then $\Xi_1[\mathbf{G}, g, g_1, h]$ consists of the set $\mathcal{M}(\mathbf{G}, 3, 3, P) \cup \{a, 1\}$ with multiplication $1x = x1 = x$ for every x, $aa = (2, g_2, 2)$,

$$a(i,k,j) = \begin{cases} (i+1, k, j), \ if \ i < 2 \\ (2, g_3g_2^{-1}k, j) = (2, g_1gk, j), \ if \ i = 2 \end{cases}$$

and

$$(i,k,j)a = \begin{cases} (i, k, j+1), \ if \ j < 2 \\ (i, kg_2^{-1}g_3, 2) = (2, kgg_1, j), \ if \ j = 2. \end{cases}$$

Multiplication within $\mathcal{M}(\mathbf{G}, 3, 3, P)$ will be as usual.

NOTE 3.3.2 *In general,*

$$\begin{aligned} g_2^{-1}g_i &= g_1^{-1}g^{-1}g_1^{-1}(g_1g)(g_1g)(g_1g)^{i-3}g_1 \\ &= g(g_1g)^{i-3}g_1 \end{aligned}$$

and likewise, $g_ig_2^{-1} = g_1(gg_1)^{i-3}g$. This means, in particular, that $a(2, g_i, 2) = (2, g_{i+1}, 2) = (2, g_i, 2)a$.

LEMMA 3.3.3 *For any centreless group* $\mathbf{G}$*, the groupoid* $\Xi_1[\mathbf{G}, g, g_1, h]$ *as constructed in Definition 3.3.1 is an INFB semigroup with* $\mathbf{B}_2^1 \notin \mathcal{V}(\Xi_1[\mathbf{G}, g, g_1, h])$.

Proof: First we check that $\Xi_1[\mathbf{G}, g, g_1, h]$ is a semigroup. Since $\mathcal{M}[\mathbf{G}, 3, 3, P]$ is a semigroup, up to symmetry we have five cases to consider.

Case 1. $a[(i,s,j)(i',t,j')] = [a(i,s,j)](i',t,j')$.
If $i \leq 1$ then the left side of this expression becomes

$$a[(i,s,j)(i',t,j')] = a(i, sP_{j,i'}t, j') = (i+1, sP_{j,i'}t, j').$$

Likewise the right side becomes

$$[a(i,s,j)](i',t,j') = (i+1,s,j)(i',t,j') = (i+1, sP_{j,i'}t, j')$$

as required. If $i = 2$ then the left side becomes

$$a[(2,s,j)(i',t,j')] = a(2, sP_{i',j}t, j') = (2, g_3 g_2^{-1} sP_{i',j}t, j')$$

and the right hand side becomes

$$[a(2,s,j)](i',t,j') = (2, g_3 g_2^{-1} s, j)(i',t,j) = (2, g_3 g_2^{-1} sP_{j,i'}t, j')$$

as required.

Case 2. $(i,s,j)[a(i',t,j')] = [(i,s,j)a](i',t,j')$
If both j and i' are less than 2 then the left side becomes

$$(i,s,j)[a(i',t,j')] = (i,s,j)(i'+1,t,j') = (i, sP_{j,i'+1}t, j')$$

and the right side becomes

$$[(i,s,j)a](i',t,j') = (i,s,j+1)(i',t,j') = (i, sP_{j+1,i'}t, j').$$

Since for any $a, b, c, d \leq 2$, $P_{a,b} = P_{c,d}$ if $a + b = c + d$, the two sides are equal.

If $j = 2$ and $i' < 2$ then the left side becomes

$$\begin{aligned}(i,s,2)[a(i',t,j')] &= (i,s,2)(i'+1,t,j') \\ &= (i,sP_{2,i'+1}t,j') \\ &= (i,sg_2^{-1}g_{i'+1+2}g_2^{-1}t,j')\end{aligned}$$

and the right side becomes

$$\begin{aligned}[(i,s,2)a](i',t,j') &= (i,sg_2^{-1}g_3,2)(i',t,j') \\ &= (i,sg_2^{-1}g_3P_{2,i'}t,j) \\ &= (i,sg_2^{-1}g_3g_2^{-1}g_{i'+2}g_2^{-1}t,j).\end{aligned}$$

To show the two sides are equal we need to show that

$$g_2^{-1}g_{i'+3}g_2^{-1} = g_2^{-1}g_3g_2^{-1}g_{i'+2}g_2^{-1}.$$

Since $g_i = (g_1g)^{i-1}g_1$ by definition, we have

$$\begin{aligned}(g_2^{-1}g_{i'+3})g_2^{-1} =& (g(g_1g)^{i'}g_1)g_1^{-1}g^{-1}g_1^{-1} \text{ (by Note 3.3.2)} \\ =& g(g_1g)^{i'-1} \\ =& g_1^{-1}(g_1g)^{i'} \\ =& g_1^{-1}(g_1g)^{i'}(g_1gg_1)(g_1gg_1)^{-1} \\ =& (g_1gg_1)^{-1}(g_1gg_1gg_1)(g_1gg_1)^{-1}(g_1g)^{i'+1}g_1(g_1gg_1)^{-1} \\ =& (g_1gg_1)^{-1}(g_1gg_1gg_1)(g_1gg_1)^{-1}g_{i'+2}(g_1gg_1)^{-1} \\ =& g_2^{-1}g_3g_2^{-1}g_{i'+2}g_2^{-1}\end{aligned}$$

as required. The proof is similar when $j < 2$ and $i' = 2$.

Finally we need to consider the case when $j = i' = 2$. In this case the left hand side becomes

$$\begin{aligned}(i,s,2)[a(2,t,j')] &= (i,s,2)(2,g_3g_2^{-1}t,j') \\ &= (i,sP_{2,2}g_3g_2^{-1}t,j') \\ &= (i,sg_2^{-1}g_4g_2^{-1}g_3g_2^{-1}t,j')\end{aligned}$$

and the right hand side becomes

$$\begin{aligned}[(i,s,2)a](2,t,j') &= (i,sg_2^{-1}g_3,2)(2,t,j') \\ &= (i,sg_2^{-1}g_3P_{2,2}t,j') \\ &= (i,sg_2^{-1}g_3g_2^{-1}g_4g_2^{-1}t,j').\end{aligned}$$

Applying the same arguments as before, we get

$$\begin{aligned}(g_2^{-1}g_4)(g_2^{-1}g_3)(g_2^{-1}) &= (gg_1gg_1)(gg_1)(g_1^{-1}g^{-1}g_1^{-1}) \text{ (by Note 3.3.2)} \\ &= gg_1g \\ &= gg_1gg_1gg_1(g_1gg_1)^{-1} \\ &= g_2^{-1}g_3g_2^{-1}g_4g_2^{-1} \text{ (by Note 3.3.2)}\end{aligned}$$

as required.

Case 3. $a[a(i,s,j)] = [aa](i,s,j)$.

If $i = 0$ then we have $a[a(0,s,j)] = a(1,s,j) = (2,s,j) = (2,g_2g_2^{-1}g_2g_2^{-1}s,j) = (2,g_2P_{2,0}s,j) = (2,g_2,2)(0,s,j) = [aa](0,s,j)$. If $i = 1$ then we have $a[a(1,s,j)] = a(2,s,j) = (2,g_3g_2^{-1}s,j) = (2,g_2g_2^{-1}g_3g_2^{-1}s,j) = (2,g_2P_{2,1}s,j) = (2,g_2,2)(1,s,j) = [aa](1,s,j)$. Finally if $i = 2$ we have $a[a(2,s,j)] = a(2,g_3g_2^{-1}s,j) = (2,(g_3g_2^{-1})^2s,j) = (2,(g_1g)^2s,j)$ (by Note 3.3.2) and $(2,(g_1g)^2s,j) = (2,g_1(gg_1)gs,j) = (2,g_4g_2^{-1}s,j)$ (again by Note 3.3.2) and $(2,g_4g_2^{-1}s,j) = (2,g_2g_2^{-1}g_4g_2^{-1}s,j) = (2,g_2P_{2,2}s,j) = (2,g_2,2)(2,s,j) = [aa](1,s,j)$ as required.

Case 4. $[a(i,s,j)]a = a[(i,s,j)a]$.

If both i and j are less than 2 then

$$[a(i,s,j)]a = (i+1,s,j)a = (i+1,s,j+1) = a(i,s,j+1) = a[(i,s,j)a].$$

If $i = 2$ and $j < 2$ then

$$[a(2,s,j)]a = (2,g_3g_2^{-1}s,j)a = (2,g_3g_2^{-1}s,j+1) = a(2,s,j+1) = a[(i,s,j)a].$$

If $i = j = 2$ we have

$$[a(2,s,2)]a = (2,g_3g_2^{-1}s,2)a = (2,g_3g_2^{-1}sg_2^{-1}g_3,2) = a(2,sg_2^{-1}g_3,2) = a[(i,s,j)a]$$

as required. The proof is similar if $i < 2$ and $j = 2$.

Case 5. $a[aa] = [aa]a$.

This follows because $a[aa] = a(2, g_2, 2) = (2, g_3 g_2^{-1} g_2, 2) = (2, g_3, 2) = (2, g_2 g_2^{-1} g_3, 2) = (2, g_2, 2)a = [aa]a$.

So $\Xi_1[\mathbf{G}, g, g_1, h]$ is a semigroup. To show it is INFB first note that $\Xi_1[\mathbf{G}, g, g_1, h]$ is a monoid and that the element a divides the idempotent $(0, e, 0)$ because

$$(0, g_2^{-1} P_{0,2}^{-1}, 0)aa(2, P_{2,2}^{-1}, 0) = (0, g_2^{-1} P_{0,2}^{-1}, 0)(2, g_2, 2)(2, P_{2,2}^{-1}, 0) = (0, e, 0).$$

However the period of $\Xi_1[\mathbf{G}, g, g_1, h]$ is d (the exponent of $\mathbf{G}$) and $a^{d+1} = (2, g_1, 2)$ so

$$(0, e, 0)a(0, e, 0) = (0, h, 0) \neq (0, g_2^{-1} g_1 g_2^{-1}, 0) = (0, e, 0)(2, g_1, 2)(0, e, 0).$$

as required by Theorem 1.1.2. Finally we need to show that $\mathbf{B}_2^1$ is not contained in the variety $\mathcal{V}(\Xi_1[\mathbf{G}, g, g_1, h])$. It is well known and easy to verify that a Rees matrix semigroup over a group of exponent d satisfies the identities $x \approx x^{d+1}$ and $(xyz)^d \approx (xz)^d$. Therefore $\mathcal{M}[\mathbf{G}, 3, 3, P]$ satisfies the identity $(xyx^2y)^d \approx (xy)^d$. Furthermore if we delete all occurrences of a given letter from this identity then the resulting identity is still satisfied by $\mathcal{M}[\mathbf{G}, 3, 3, P]$. Therefore the monoid obtained from $\mathcal{M}[\mathbf{G}, 3, 3, P]$ by adjoining an identity element satisfies $(xyx^2y)^d \approx (xy)^d$. So in order to show that $\Xi_1[\mathbf{G}, g, g_1, h]$ satisfies $(xyx^2y)^d \approx (xy)^d$ we need only check cases where the element a is assigned to at least one of the letters x and y. If a is assigned to both x and y or if a is assigned to just one of these and 1 is assigned to the other then both sides simply equal a^d. If a is assigned to x but (i, s, j) is assigned to y, then xy becomes (i', t, j) for some i' and some $t \in \mathbf{G}$. In this case, both sides of the identity become the idempotent in the subgroup $H_{i',j}$ of all elements of the form (i', r, j), where $r \in \mathbf{G}$. The case when a is assigned to y and (i, s, j) is assigned to x is similar. Thus

$$\Xi_1[\mathbf{G}, g, g_1, h] \models (xyx^2y)^d \approx (xy)^d.$$

However $\mathbf{B}_2^1 \not\models (xyx^2y)^d \approx (xy)^d$ since (as noted in the introduction to this chapter) $\mathbf{B}_2^1$ contains two elements, a and b, such that ab is a nonzero idempotent but $a^2 = 0$. Since the left side of $(xyx^2y)^d \approx (xy)^d$ contains x^2 but the right side is of the form $(xy)^d$, assigning a to x and b to y ensures the left side becomes 0 but the right side becomes the nonzero idempotent. □

We will say that $\Xi_1[\mathbf{G}, g, g_1, h]$ is a *small INFB finite semigroup of the first kind* and denote the set of all such monoids by Ξ_1.

NOTE 3.3.4 *In [74] a finite INFB monoid* $\mathbf{T}$ *is presented for any centreless group* $\mathbf{G}$ *with a non identity element* g *with the property* $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{T})$. *By letting* e *be the identity element of* $\mathbf{G}$ *it is possible to show that the monoid* $\Xi_1[\mathbf{G}, g, g^{-1}, e]$ *is a (proper) homomorphic image of* $\mathbf{T}$.

Note also that if $\mathbf{S_3}$ is the six element centreless group then $\Xi_1[\mathbf{S_3}, g, g_1, h]$ has exactly 56 elements for any valid choice of g, g_1 and h from $\mathbf{S_3}$. By Corollary 3.2.12 this is the smallest possible size for such a semigroup.

For integers a, b, r we will use the notation $a + (b \ mod(r))$ to denote the sum of a with the smallest non-negative element of the equivalence class $b \ mod(r)$. We will also use the notation $[a/b]$ to denote the *integer part* of the rational number a/b. For example, for any pair of integers n and m we have $n = m[n/m] + (n \ mod(m))$.

DEFINITION 3.3.5 *Let* $\mathbf{G}$ *be a centreless group with exponent* d *and identity element* e *and let* $\langle a \rangle$ *be a finite cyclic semigroup of index* 2 *and period* p *generated by an element* a. *Suppose* p *has two divisors* l *and* r, *not both* 1, *such that there are elements* L *and* R *of* $\mathbf{G}$ *with order* p/l *and* p/r *respectively and a mapping* $f : \langle a \rangle \to \mathbf{G}$ *satisfying:*

(i) $f(a) \neq f(a^{1+p})$,

(ii) for all $i, j \geq 0$ *with* $i + j \leq 1 + p$,

$$f(a^{2+i+j}) = L^{[j/l]} f(a^{2+i+(j \ mod(l))}) = f(a^{2+i+(j \ mod(r))}) R^{[j/r]},$$

(iii) for any $j \leq p$, if $f(a^{i+j}) = f(a^i)$ for every i with $1+p \geq i \geq 2$ then $p = j$.
Then $\Xi_2[\mathbf{G}, L, R, f, p]$ is the groupoid

$$\mathcal{M}(\mathbf{G}, 2+l, 2+r, P) \cup \{1, a, a^2, \ldots, a^{1+p}\}$$

where $P_{i,j} = f(a^{i+j})$, $P_{0,0} = e$ and multiplication is defined by

$$a(i,k,j) = \begin{cases} (i+1,k,j), \ if \ i < 2, \\ (2+((i-1)mod(r)), R^{[(i-1)/r]}k, j), \ if \ 2 \leq i \leq 2+r \end{cases}$$

$$(i,k,j)a = \begin{cases} (i,k,j+1), \ if \ j < 2, \\ (i, kL^{[(j-1)/l]}, 2+((j-1)mod(l))), \ if \ 2 \leq j \leq 2+l \end{cases}$$

and $a^i x = a^{i-1}(ax)$, $xa^i = (xa)a^{i-1}$.

LEMMA 3.3.6 *In general, $a^n(i,s,j) = (2+((i+n-2)mod(r)), R^{[(i+n-2)/r]}s, j)$ and $(i,s,j)a^n = (i, sL^{[(j+n-2)/l]}, 2+((j+n-2)mod(l)))$.*

Proof: We use induction. Firstly $a(i,s,j) = (2+((i-1)mod(r)), R^{[(i-1)/r]}s, j)$ so the claim is true for $n = 1$. Assume that

$$a^n(i,s,j) = (2+((i+n-2)mod(r)), R^{[(i+n-2)/r]}s, j).$$

We now show that $a^{n+1}(i,s,j) = (2+((i+n-1)mod(r)), R^{[(i+n-1)/r]}s, j)$.

Now

$$\begin{aligned} & a^{n+1}(i,s,j) \\ =& a(2+((i+n-2)mod(r)), R^{[(i+n-2)/r]}s, j), \text{ (by assumption)} \\ =& (2+(2+((i+n-2)mod(r))-1)mod(r)), \\ & \qquad R^{[(2+((i+n-2)mod(r))-1)/r]}R^{[(i+n-2)/r]}s, j) \\ =& (2+((i+n-1)mod(r)), R^{[(1+((i+n-2)mod(r)))/r]}R^{[(i+n-2)/r]}s, j). \end{aligned}$$

It remains to show that $R^{[(1+((i+n-2)mod(r)))/r]}R^{[(i+n-2)/r]} = R^{[(i+n-1)/r]}$. Let t be the element $R^{[(1+((i+n-2)mod(r)))/r]}R^{[(i+n-2)/r]} \in \mathbf{G}$. Now either $1+((i+n-2)mod(r)) < r$

or $1+((i+n-2)mod(r)) = r$. If $1+((i+n-2)mod(r)) < r$ then $((i+n-1)mod(r)) < r$ and so $[\frac{i+n-2}{r}] = [\frac{i+n-1}{r}]$. In this case $t = R^{[(i+n-2)/r]} = R^{[(i+n-1)/r]}$ as required. If $1+((i+n-2)mod(r)) = r$ then $R^{[(1+((i+n-2)mod(r)))/r]} = R$, $((i+n-2)mod(r)) = r-1$ and $[\frac{i+n-2}{r}] = [\frac{i+n-1}{r}]-1$. So $t = RR^{[(i+n-1)/r]-1} = R^{[(i+n-1)/r]}$ as required. Therefore by induction the result is true for all $n \geq 1$.

The corresponding result for multiplication on the right follows by symmetry. □

LEMMA 3.3.7 *The groupoid $\Xi_2[\mathbf{G}, L, R, f, p]$ as constructed in Definition 3.3.5 is an INFB semigroup and $\mathbf{B}_2^1 \notin \mathcal{V}(\Xi_2[\mathbf{G}, L, R, f, p])$.*

Proof: First we will show that $\Xi_2[\mathbf{G}, L, R, f, p]$ is a semigroup. Since $\mathcal{M}(\mathbf{G}, 2+r, 2+l, P)$ is a semigroup we have, up to symmetry, four cases to consider.

Case 1. $a^n[(i,s,j)(i',t,j')] = [a^n(i,s,j)](i',h,j')$.

This is similar to Case 1 in Lemma 3.3.3: multiplying an element (i,s,j) on the left by a^n gives an element of the form (k, rs, j) where r is some element of $\mathbf{G}$, k is a number and both r and k depend only on the numbers n and i. Thus

$$a^n[(i,s,j)(i',t,j')] = a^n(i, sP_{j,i'}t, j') = (k, rsP_{j,i'}t, j')$$

and

$$[a^n(i,s,j)](i',h,j') = [a^n(i,s,j)](i',h,j') = (k,rs,j)(i',h,j') = (k, rsP_{j,i'}t, j')$$

as required.

Case 2. $(i,s,j)[a^n(i',t,j')] = [(i,s,j)a^n](i',t,j')$.

The case when j, n and i' are sufficiently small that both $i'+n$ and $j+n$ are less than or equal to 2 is essentially the same as the first case considered in Case 2 of Lemma 3.3.3. Now say that $i'+n \leq 2$ but $j+n > 2$(if $i'+n \geq 2$ and $j+n < 2$ then the proof is similar). In this case the left side becomes

$$(i,s,j)[a^n(i',t,j')] = (i,s,j)(i'+n,t,j') = (i, sP_{j,i'+n}t, j').$$

Now using Lemma 3.3.6 the right hand side becomes

$$\begin{aligned}[(i,s,j)a^n](i',t,j') &= (i, sL^{[(j+n-2)/l]}, 2+((j+n-2)mod(l)))(i',t,j') \\ &= (i, sL^{[(j+n-2)/l]}P_{2+((j+n-2)mod(l)),i'}t, j').\end{aligned}$$

If associativity is to hold then

$$L^{[(j+n-2)/l]}P_{i',2+((j+n-2)mod(l))} = P_{j,i'+n}.$$

The left side of this is

$$\begin{aligned}L^{[(j+n-2)/l]}P_{i',2+((j+n-2)mod(l))} &= L^{[(j+n-2)/l]}f(a^{2+i'+((j+n-2)mod(l))}) \\ &= f(a^{i'+j+n}) \\ &= P_{j,i'+n}\end{aligned}$$

which is the right hand side as required.

Finally consider the case when $i'+n$ and $j+n$ are both greater than 2. In this case the left side becomes

$$\begin{aligned}(i,s,j)[a^n(i',t,j')] =&(i,s,j)[a^n(2+i'-2,t,j')] \\ =&(i,s,j)(2+((n+i'-2)mod(r)), R^{[(n+i'-2)/r]}t, j') \\ &\text{(by Lemma 3.3.6)} \\ =&(i, sP_{j,2+((n+i'-2)mod(r))}R^{[(n+i'-2)/r]}t, j')\end{aligned}$$

and the right side becomes

$$\begin{aligned}[(i,s,j)a^n](i',t,j') &= (i, sL^{[(j+n-2)/l]}, 2+((j+n-2)mod(l)))(i',t,j') \\ &= (i, sL^{[(j+n-2)/l]}P_{2+((j+n-2)mod(l))),i'}t, j').\end{aligned}$$

Now

$$\begin{aligned} P_{j,2+((n+i'-2)mod(r))}R^{[(n+i'-2)/r]} &= f(a^{2+((n+i'-2)mod(r))+j})R^{[(n+i'-2)/r]} \\ &= f(a^{n+i'+j}) \\ &= L^{[(j+n-2)/l]}f(a^{2+i'+((j+n-2)mod(l))}) \\ &= L^{[(j+n-2)/l]}P_{2+((j+n-2)mod(l))),i'} \end{aligned}$$

as required.

Case 3. $a^n[a^m(i,t,j)] = [a^n a^m](i,h,j)$.

This follows immediately since from Definition 3.3.5 we have that $a^n(i,h,j) = a^{n-1}(a(i,h,j))$.

Case 4. $[a^n(i,s,j)]a^m = a^n[(i,s,j)a^m]$.

This follows for essentially the same reasons as the result in Case 1.

Therefore associativity holds and $\Xi_2[\mathbf{G}, L, R, f, p]$ is a semigroup. Also, by Theorem 1.1.2, INFB occurs at the dividing pair $(a,(0,e,0))$ since

$$(0,e,0)a(0,e,0) = (0,f(a),0)$$

and by Lemma 3.3.6,

$$(0,e,0)a^{d+1}(0,e,0) = (0,e,0)a^{p+1}(0,e,0) = (0,f(a^{p+1}),0)$$

where $f(a) \neq f(a^{p+1})$.

Finally, to show that $\mathbf{B}_2^1$ is not contained in $\mathcal{V}(\Xi_2[\mathbf{G}, L, R, f, p])$ we again use the identity $(xyx^2y)^q \approx (xy)^q$ where q is the period of $\Xi_2[\mathbf{G}, L, R, f, p]$, or equivalently the lowest common multiple of the exponent d of $\mathbf{G}$ and the period p of $\langle a \rangle$. Since both p and d divide q as numbers, it follows that $\Xi_2[\mathbf{G}, L, R, f, p]$ satisfies this identity for essentially the same reasons as in the proof of Lemma 3.3.3. □

We will say that $\Xi_2[\mathbf{G}, L, R, f, p]$ is a *small INFB semigroup of the second kind* and denote the set of all such monoids by Ξ_2.

We will now construct an example.

it follows that the direct product of $\mathbf{L}^1$ with any finite group also generates such a variety; see Corollary 3.1.5 below).

Summarising and combining the ideas above we obtain the following theorem.

THEOREM 4.1.13 *(i) For any semigroup* $\mathbf{S}_1$ *(finite or otherwise) there are finite semigroups* $\mathbf{S}_2$ *and* $\mathbf{S}_3$ *generating hereditarily finitely based varieties so that* $\mathbf{S}_1 \times \mathbf{S}_2 \times \mathbf{S}_3$ *generates a variety with uncountably many subvarieties.*
(ii) If $\mathbf{M}$ *is a monoid of index more than two then there is a finite group* $\mathbf{G}$ *generating a hereditarily finitely based variety with only* 3 *subvarieties so that* $\mathbf{M} \times \mathbf{G}$ *generates a variety with uncountably many subvarieties.*
(iii) If $\mathbf{M}$ *is a monoid of index less than or equal to two then either* $\mathbf{M}$ *satisfies both* $xyx \approx xxy$ *and* $xyx \approx yxx$ *or there is a finite semigroup* $\mathbf{S}$ *generating a hereditarily finitely based variety so that* $\mathbf{M} \times \mathbf{S}$ *generates a variety with uncountably many subvarieties.*

Proof: (i) For $\mathbf{S}_2$ and $\mathbf{S}_3$ one can take, for example, the semigroups $\mathbf{L}^1$ and $S(\{aab\})$ or the semigroups $\mathbf{B}$ and $S(\{aa\})$.

(ii) The monoid $S(\{aa\})$ is contained in the variety generated by $\mathbf{M}$ and therefore the claim follows by taking $\mathbf{G}$ to be the group $\mathbf{B}$ above. To obtain a aperiodic example one may replace the group $\mathbf{B}$ in this argument by the direct product of $\mathbf{L}^1$ with its right dual $\mathbf{R}^1$ and obtain a similar result. The semigroup $\mathbf{L}^1 \times \mathbf{R}^1$ generates a band variety with a lattice of subvarieties consisting of 13 elements.

(iii) If $\mathbf{M}$ does not satisfy one of the described identities then one of the semigroups $\mathbf{M} \times S(\{aab\})$ or $\mathbf{M} \times S(\{abb\})$ generates a variety whose identities are closed under deletion, have index three and do not contain either of the identities $xyx \approx xxy$ and $xyx \approx yxx$. By the last part of Theorem 4.1.2, one of these semigroups generates a variety with uncountably many subvarieties. □

In connection with part (iii) of this theorem we note that a monoid of index one satisfying both $xyx \approx xxy$ and $xyx \approx yxx$ is a semilattice of groups (a Clifford

EXAMPLE 3.3.8 *As our group we will take the symmetric group* $\mathbf{S_3}$ *of order* 6 *(and exponent* 6*) with presentation* $\langle L, R; L^3 = R^2 = e, LR = RL^2 \rangle$. *Let* p *be the number* 6. *The orders of* L *and* R *are* 3 *and* 2 *respectively so the numbers* l *and* r *required by Definition 3.3.5 are* 2 *and* 3 *respectively. Finally we define our mapping* f *according to the following table:*

$f(a)$	$f(a^2)$	$f(a^3)$	$f(a^4)$	$f(a^5)$	$f(a^6)$	$f(a^7)$
L	L	R	L^2	LR	$L^3 = R^2 = e$	L^2R

It is easily verified that f *satisfies the requirements of Definition 3.3.5. So the sandwich matrix* P *of the completely simple ideal of* $\Xi_2[\mathbf{S_3}, L, R, f, p]$ *is*

$$\begin{pmatrix} e & L & L & R \\ L & L & R & L^2 \\ L & R & L^2 & LR \\ R & L^2 & LR & e \\ L^2 & LR & e & L^2R \end{pmatrix}$$

We now show that the class $\Xi_1 \cup \Xi_2 \cup \mathbf{B_2^1}$ contains all minimal finite INFB semigroups.

THEOREM 3.3.9 *Let* $\mathbf{S}$ *be a finite semigroup. The* $\mathbf{S}$ *is INFB if and only if there is a monoid* $\mathbf{T} \in \Xi_1 \cup \Xi_2 \cup \{\mathbf{B_2^1}\}$ *with* $\mathbf{T} \in \mathcal{V}(\mathbf{S})$.

Proof: The "if" implication follows immediately from the property of being INFB. Now we show that the reverse implication is also true. Firstly by Theorem 1.1.2 we may assume that $\mathbf{S}$ is a monoid with identity element 1 and that there is an idempotent e and an element a so that INFB occurs at (a, e). Now assume that $\mathbf{B_2^1} \notin \mathcal{V}(\mathbf{S})$. So by Lemma 3.1.1, $ea^ie \in \mathbf{S}_e$ for every $i \geq 0$. We will take a series of subsemigroups and homomorphic images until we arrive at a semigroup isomorphic

to one from $\Xi_1 \cup \Xi_2$. This process is equivalent to taking a single homomorphic image of a subsemigroup of $\mathbf{S}$ (see [66] for example). The small INFB semigroup we arrive at is therefore a *divisor* of $\mathbf{S}$.

Consider the subsemigroup $\mathbf{T}$ of $\mathbf{S}$ generated by the set $\mathbf{S}_e \cup \{a, 1\}$. Now a still divides e in $\mathbf{T}$ since $(e)a(e(eae)^{-1}) = e$. Let i and p be the index and period respectively of $\langle a \rangle$ (the subsemigroup generated by a). By Lemma 3.2.1, i is at least 2. Also since p divides the period of $\mathbf{T}$ and the period of $\mathbf{T}$ divides the period of $\mathbf{S}$ (say d), the property $eae \not\equiv ea^{p+1}e$ modulo $\Gamma(\mathbf{S}_e)$ is preserved and therefore $\mathbf{T}$ is an INFB submonoid of $\mathbf{S}$. Note that $\mathbf{T}_e$ is identical to $\mathbf{S}_e$.

Since $\mathbf{T}$ is generated by $\mathbf{T}_e \cup \{a, 1\}$ and $ea^k e \in \mathbf{T}_e$ for every $k \geq 0$ (recall $a^0 = 1$), every element in $\mathbf{T}$ except 1 can be considered as a word of the form $a^n s a^m$ where n and m are non negative integers and $s \in \mathbf{T}_e$. We now want to replace the non-nilpotent group $\mathbf{T}_e$ with a centreless (and therefore also non-nilpotent) group. Consider the equivalence θ_1 defined as

$$\{(x, y) : x = y \text{ or } x = a^n s a^m,\ y = a^n t a^m,\ n, m \geq 0 \text{ and } s \equiv t \ mod\ \Gamma(\mathbf{T}_e)\}.$$

This is a congruence since if $a^n s a^m$ and $a^n t a^m$ are equivalent modulo θ_1 then

$$a^{n'} g a^{m'} a^n s a^m = a^{n'} g e a^{m'+n} e s a^m$$

and

$$a^{n'} g a^{m'} a^n t a^m = a^{n'} g e a^{m'+n} e t a^m$$

for any non-negative integers n' and m' and $g \in \mathbf{T}_e$. Since $s \equiv t \ mod\ \Gamma(\mathbf{T}_e)$ we must have

$$g e a^{m'+n} e s \equiv g e a^{m'+n} e t \ mod\ \Gamma(\mathbf{T}_e)$$

and therefore

$$(a^{n'} g e a^{m'+n} e s a^m, a^{n'} g e a^{m'+n} e t a^m) \in \theta_1.$$

So θ_1 is a left congruence and likewise, by symmetry, a right congruence. Let $\bar{\mathbf{T}}$ denote the monoid $\mathbf{T}/\theta_1$. This is still an INFB monoid since eae and $ea^{d+1}e$ were

not equivalent modulo $\Gamma(\mathbf{T}_e)$ so

$$(eae)/\theta_1 = (eae)\Gamma(\mathbf{T}_e) \neq (ea^{d+1}e)\Gamma(\mathbf{T}_e) = (ea^{d+1}e)/\theta_1.$$

To avoid unnecessarily complicated expressions we will relabel the equivalence classes of $\bar{\mathbf{T}}$ so that a/θ_1 becomes a and e/θ_1 becomes e. By the definition of the upper central series, the group $\bar{\mathbf{T}}_e$ is centreless and so $\Gamma(\bar{\mathbf{T}}_e) = \{e\}$. Therefore two elements of $\bar{\mathbf{T}}_e$ are equivalent modulo $\Gamma(\bar{\mathbf{T}}_e)$ if and only if they are equal.

Let j be the smallest positive integer so that $ea^{i-j}e \neq ea^{i+p-j}e$ (recall that i is the index of $\langle a \rangle$). Such a j exists since $eae \neq ea^{p+1}e$ and $ea^{i-(i-1)}e = eae$ and $ea^{i+p-(i-1)}e = ea^{p+1}e$. So by the choice of j, for any $k < j$, $ea^{i-k}e = ea^{i+p-k}e$. Now a divides $a^{i-j-1}(ea^{i-j-1}e)^{-1}$ since

$$a^{i-j-1}(ea^{i-j-1}e)^{-1} = a^{i-j-1}(ea^{i-j-1}e)^{-1} \times a \times (eae)^{-1}.$$

Also $a^{i-j-1}(ea^{i-j-1}e)^{-1}$ is idempotent since

$$\begin{aligned} & a^{i-j-1}(ea^{i-j-1}e)^{-1}a^{i-j-1}(ea^{i-j-1}e)^{-1} \\ &= a^{i-j-1}(ea^{i-j-1}e)^{-1}ea^{i-j-1}e(ea^{i-j-1}e)^{-1} \\ &= a^{i-j-1}(ea^{i-j-1}e)^{-1}. \end{aligned}$$

Therefore $(a, a^{i-j-1}(ea^{i-j-1}e)^{-1})$ is a dividing pair and the set $\{a^{i-j-1}s : s \in \bar{\mathbf{T}}_e\}$ is a subgroup of $\bar{\mathbf{T}}$ isomorphic to $\bar{\mathbf{T}}_e$ (it is easily verified that the map $f : \bar{\mathbf{T}}_e \to \{a^{i-j-1}s : s \in \bar{\mathbf{T}}_e\}$ given by $f(s) = a^{i-j-1}(ea^{i-j-1}e)^{-1}s$ is an isomorphism). Now

$$(a^{i-j-1}(ea^{i-j-1}e)^{-1})a(a^{i-j-1}(ea^{i-j-1}e)^{-1})$$

is equal to

$$a^{i-j-1}(ea^{i-j-1}e)^{-1}ea^{i-j}e(ea^{i-j-1}e)^{-1}$$

and

$$(a^{i-j-1}(ea^{i-j-1}e)^{-1})a^{p+1}(a^{i-j-1}(ea^{i-j-1}e)^{-1})$$

is equal to

$$a^{i-j-1}(ea^{i-j-1}e)^{-1}ea^{i+p-j}e(ea^{i-j-1}e)^{-1}.$$

But since $ea^{i-j}e$ and $ea^{i+p-j}e$ are not equal, neither can be

$$a^{i-j-1}(ea^{i-j-1}e)^{-1}ea^{i-j}e(ea^{i-j-1}e)^{-1}$$

and

$$a^{i-j-1}(ea^{i-j-1}e)^{-1}ea^{i+p-j}e(ea^{i-j-1}e)^{-1}.$$

Therefore INFB occurs at $(a, a^{i-j-1}(ea^{i-j-1}e)^{-1})$ and for every $k > 1$,

$$(a^{i-j-1}(ea^{i-j-1}e)^{-1})a^k(a^{i-j-1}(ea^{i-j-1}e)^{-1})$$

is equal to

$$(a^{i-j-1}(ea^{i-j-1}e)^{-1})a^{p+k}(a^{i-j-1}(ea^{i-j-1}e)^{-1}).$$

Let the idempotent $a^{i-j-1}(ea^{i-j-1}e)^{-1}$ be denoted by f and let $\mathbf{U}$ be the submonoid of $\bar{\mathbf{T}}$ generated by $\bar{\mathbf{T}}_f \cup \{a, 1\}$. Using the same kind of argument as was used in the case of $\mathbf{T}$, we have that $\mathbf{U}$ is an INFB submonoid of $\bar{\mathbf{T}}$ and INFB occurs at (a, f). However, as was noted above, $fa^kf = fa^{k+p}f$ in $\mathbf{U}$ for all $k > 1$.

Now consider the equivalence on the set $\{a, a^2, a^3, \ldots a^{i+p-1}\}$ given by

$$\phi_p = \{(a^j, a^k) : j = k \text{ or } j, k > 1 \text{ and } j \equiv k(mod(p))\}.$$

Since $sfa^kft = sfa^{k+p}ft$ in $\mathbf{U}$ for all $k > 1$ and any $s, t \in \mathbf{U}_f$, ϕ_p generates a congruence θ_2 on $\mathbf{U}$ equal to

$$\{(x, y) : x = y \text{ or } (x, y) \in \phi_p;$$
$$\text{or } x = a^{j_1}sa^{k_1},\ y = a^{j_2}sa^{k_2} \text{ and both } (a^{j_1}, a^{j_2}), (a^{k_1}, a^{k_2}) \in \phi_p\}.$$

Let $\bar{\mathbf{U}}$ be the semigroup $\mathbf{U}/\theta_2$. For the sake of simplicity we will relabel the equivalence classes so that a/θ_2 becomes a and f/θ_2 becomes e. So (a, e) is a dividing pair, $\bar{\mathbf{U}}_e$ is centreless, $ea^je \in \bar{\mathbf{U}}_e$ for all $j \geq 0$, and $eae \neq ea^{1+p}e$. Furthermore, the index of $\langle a \rangle$ (the subsemigroup generated by a) is now 2, that is $a^2 = a^{2+p}$.

Consider now the subsemigroup

$$\mathbf{C} = \{a^i s a^j : i, j \geq 0, s \in \bar{\mathbf{U}}_e\}$$

This is an ideal of the semigroup $\bar{\mathbf{U}}$ and every element in $\mathbf{C}$ divides every other element since for any $i, i', j, j' \geq 0$, $s, t \in \bar{\mathbf{U}}_e$,

$$a^i s a^j = (a^i t^{-1}(e a^{i'} e)^{-1})(a^{i'} t a^{j'})((e a^{j'} e)^{-1} s a^j).$$

Thus $\mathbf{C}$ is a completely simple subsemigroup of $\bar{\mathbf{U}}$. It is clear also that the $\mathcal{H}$-classes of $\mathbf{C}$ are sets of the form $\{a^i s a^j : s \in \bar{\mathbf{U}}_e\}$ and we will denote such an $\mathcal{H}$-class by $H_{i,j}$. The proof will now split into two cases. The first is the situation when a^2 is contained in $\mathbf{C}$. The corresponding semigroups will be elements of Ξ_1, the INFB semigroups of the first kind. The second situation is when $a^2 \notin \mathbf{C}$. In this case it is possible that some further reduction may be made.

Case 1. $a^2 \in \mathbf{C}$.

In this case the set $\{a^2, a^3, \ldots a^{1+p}\}$ is a cyclic subgroup of some group $H_{i,j}$. Now by the Rees-Suschkewitz theorem, $\mathbf{C}$ is isomorphic to a Rees Matrix Semigroup over the centreless group $\bar{\mathbf{U}}_e$ with sandwich matrix P. Since every element in $\mathbf{C}$ is of the form $a^i s a^j$ and $a^2 \in \mathbf{C}$, P must be at most a 3×3 matrix. Since Theorem 3.2.11 shows that the sets $\{1, a\}$ and each $H_{i,j}$ for $i, j \leq 2$ are disjoint, P must be exactly a 3×3 matrix. Now if $a^2 = a^i s a^j$ where $i < 2$ and $j \leq 2$ then every element in $\mathbf{C}$ can be written in the form $a^{i'} t a^{j'}$ for $t \in \bar{\mathbf{U}}_e$, $i' < 2$ and $j' \leq 2$ and then P is be only an $i' \times 3$ or $3 \times i'$ matrix, a contradiction. Therefore, by symmetry, $a^2 = a^2 s a^2$ for some $s \in \bar{\mathbf{U}}_e$ and the subgroup $\{a^2, a^3, \ldots a^{1+p}\}$ is a cyclic subgroup of $H_{2,2}$. Note also that $\{a^2, a^3, \ldots a^{1+p}\}$ is generated by a^{1+p} since $(a^{1+p})^n = a^{n+np} = a^n$. For some $g_1 \in \bar{\mathbf{U}}_e$, $a^{1+p} = a^2 g_1 a^2$. Let g be the element $e a^4 e$ and define a map $\iota : \bar{\mathbf{U}} \to \Xi_1(\bar{\mathbf{U}}_e, g, g_1, eae)$ by

$$\iota(1) = 1,\ \iota(a) = a, \iota(a^i s a^j) = (i, s, j).$$

We show that ι is an isomorphism. It is certainly a bijection since during the arguments above we have shown that $\bar{\mathbf{U}}$ contains only elements of the form 1, a,

and a^isa^j (for $i,j \leq 2$ and $s \in \bar{\mathbf{U}}_e$) and Theorem 3.2.11 shows that these are distinct in any finite INFB semigroup whose variety does not contain $\mathbf{B}_2^1$. We need to show that for any elements $x, y \in \bar{\mathbf{U}}$, $\iota(xy) = \iota(x)\iota(y)$. The case when x or y is 1 is trivial.

Consider the case when $x = a^isa^j$ and $y = a^{i'}ta^{j'}$. Now

$$\iota(a^isa^ja^{i'}ta^{j'}) = \iota(a^isea^{i'+j}eta^{j'}) = (i, sea^{i'+j}et, j')$$

and

$$\iota(a^isa^j)\iota(a^{i'}ta^{j'}) = (i,s,j)(i',t,j') = (i, sP_{j,i'}t, j').$$

As in Definition 3.3.1, put $g_i = (g_1g)^{i-1}g_1$. If $i' + j \geq 2$ then

$$a^{j+i'} = (a^{1+p})^{j+i'} = (a^2g_1a^2)^{j+i'} = a^2(g_1ea^4e)^{j+i'-1}g_1a^2 = a^2g_{j+i'}a^2$$

and therefore $ea^{j+i'}e = ea^2eg_{j+i'}ea^2e$. Now $a^2 = a^2g_2a^2$ so $ea^2e = ea^2g_2a^2e$. Therefore $g_2 = (ea^2e)^{-1}$. This implies that

$$ea^{j+i'}e = ea^2eg_{j+i'}ea^2e = g_2^{-1}g_{j+i'}g_2^{-1} = P_{j,i'}$$

as required. If $i' = j = 0$ then

$$P_{j,i'} = P_{0,0} = e = ea^0e$$

as required. Finally if $i' = 1$ and $j = 0$ (the case when $i' = 0$ and $j = 1$ follows by symmetry) then

$$\iota(a^isata^{j'}) = (i, seaet, j') = (i, sP_{0,1}t, j') = (i,s,0)(1,t,j') = \iota(a^is)\iota(ata^{j'}),$$

also as required.

Now consider the case when $x = a$ and $y = a^isa^j$. Firstly assume $i = 0$. Then

$$\iota(aa^isa^j) = \iota(asa^j) = (1,s,j) = a(0,s,j) = \iota(a)\iota(sa^j),$$

as required.

Now assume that $i > 0$. Therefore

$$\iota(aa^isa^j) = \iota(a^{i+1}sa^j) = \iota(a^2g_{i+1}a^2sa^j) = (2, g_{i+1}ea^2es, j).$$

But $ea^2e = g_2^{-1}$, so $\iota(aa^isa^j) = (2, g_{i+1}g_2^{-1}s, j)$. If $i = 1$ then

$$(2, g_{i+1}g_2^{-1}s, j) = (2, s, j) = a(1, s, j) = \iota(a)\iota(asa^j)$$

as required. If $i = 2$ then

$$(2, g_{i+1}g_2^{-1}s, j) = (2, g_3g_2^{-1}s, j) = a(2, s, j) = \iota(a)\iota(a^2sa^j),$$

also as required.

Up to symmetry, the only remaining case is when $x = y = a$. That $\iota(aa) = \iota(a)\iota(a)$ follows immediately since in $\bar{\mathbf{U}}$ we have $a^2 = a^2g_2a^2$ while $a^2 = (2, g_2, 2)$ in $\Xi_1(\bar{\mathbf{U}}_e, g, g_1, eae)$. Therefore ι is an isomorphism.

Case 2. $a^2 \notin \mathbf{C}$.

Since $\{a^2, a^3, \ldots, a^{1+p}\}$ forms a cyclic subgroup of $\bar{\mathbf{U}}$, it must be that $a^i \notin \mathbf{C}$ for all $i \geq 0$. Recall that if $i > 1$ then $ea^ie = ea^{i+p}e$ and that ϕ_p is the equivalence

$$\{(a^j, a^k) : j = k \text{ or } j, k > 1 \text{ and } j \equiv k(mod(p))\}$$

on the set $\{a, a^2, a^3, \ldots\}$. Let q be the smallest number such that for all $i > 1$, $ea^ie = ea^{i+q}e$. It is easily verified that the equivalence θ_3 given by

$$\{(x, y) : x = y; \text{ or } (x, y) \in \phi_q;$$
$$\text{or } x = a^isa^j, \ y = a^{i'}sa^{j'}, \text{ and both } (a^i, a^{i'}), (a^j, a^{j'}) \in \phi_q\}$$

is a congruence that preserves the property of being an INFB monoid. Let the semigroup $\bar{\mathbf{U}}/\theta_3$ be denoted $\mathbf{V}$ and let the equivalence classes a/θ_3 and e/θ_3 be relabeled a and e respectively. Note that the group $\mathbf{V}_e = \bar{\mathbf{U}}_e/\theta_3$ is isomorphic to $\bar{\mathbf{U}}_e$ and that the period of $\langle a \rangle$ (the subsemigroup of $\mathbf{V}$ generated by a) is now the number q.

In $\mathbf{V}$, $ea^ie = ea^{i+j}e$ for all $i > 1$ if and only if $j = q$. If $ea^2 = ea^{2+j}$ then $ea^2a^{i-2} = ea^i = ea^{2+j}a^{i-2} = ea^{i+j}$ for any $i \geq 2$. Therefore the number j in these equations must be q. Likewise $a^2e = a^{2+j}e$ if and only if $j = q$. Say there exists $j \geq 2$ such that $ea^j = ga$ for some $g \in \mathbf{V}_e$. Then $ea^ja^q = ea^j$ and $gaa^q = ga^{q+1}$. Therefore $gae = ea^je = ga^{q+1}e$, a contradiction since $eae \neq ea^{q+1}e$. It follows that no such j exists and likewise that there is no integer $j \geq 2$ such that $a^je = ag$. This also guarantees that no $j \geq 1$ exists so that $a^je = g$ or $ea^j = g$ since then, for example, $a^{j+1}e = ag$. Now let r be the smallest number such that $a^2e = a^{2+r}R$ for some $R \in \mathbf{V}_e$ and l be the smallest number such that $ea^2 = La^{2+l}$ for some $L \in \mathbf{V}_e$. Now r must divide q since otherwise there are numbers k and k' such that $kr \equiv k'(mod(q))$ and $k' < r$. In this case $a^{2+kr}e = a^2R^k$ and $a^{2+kr}e = a^{2+k'}e$, contradicting the minimality of r. Likewise, l must divide q also: say $q = nr = ml$. Now since $a^{2+q}e = a^2e$ and $a^{2+q}e = a^{2+nr}e = a^2R^n$ we must have that $R^n = e$ and therefore the order of R divides q. Let the order of R be k (note that k necessarily divides n). Then $a^{2+rk}e = a^2R^k = a^2e$ and therefore $ea^ie = ea^{i+rk}e$ for every $i > 1$. By the choice of $\mathbf{V}$ however, this is true only if $rk = q$. Therefore the order of R is n and, by symmetry, the order of L is m.

If $l = r = 1$ then L and R have the same order and for any integer $i \geq 0$, $L^i(ea^2e) = ea^{2+i}e = (ea^2e)R^i$. Furthermore for any $k \geq 1$ and $i, j \geq 0$,

$$a^{2+k} \times a^isa^j = a^2R^{i+k}sa^j$$

and

$$\begin{aligned} a^2R^k(ea^2e)^{-1}a^2 \times a^isa^j &= a^2R^k(ea^2e)^{-1}ea^2R^isa^j \\ &= a^2R^{i+k}sa^j. \end{aligned}$$

Likewise $a^isa^j \times a^{2+k} = a^isa^j \times a^2(ea^2e)^{-1}L^ka^2$. But $ea^2eR^k = L^kea^2e$ and so

$$a^2(ea^2e)^{-1}L^ka^2 = a^2R^k(ea^2e)^{-1}a^2.$$

Therefore multiplication on the left of an element of $\mathbf{C}$ by a^{2+k} is the same as multiplication on the left by $a^2R^k(ea^2e)^{-1}a^2$ (or equivalently by $a^2(ea^2e)^{-1}L^ka^2$) and likewise with multiplication on the right. We also have

$$a \times a^{2+k} = a^{2+(k+1)} = a^{2+k} \times a$$

and

$$a \times a^2R^k(ea^2e)^{-1}a^2 = a^2R^{k+1}(ea^2e)^{-1}a^2 = a^2R^k(ea^2e)^{-1}a^2 \times a.$$

Therefore the equivalence θ_4 given by

$$\{(x,y) : x = y; \text{ or } x = a^{2+k} \text{ and}$$
$$y = a^2R^k(ea^2e)^{-1}a^2; \text{ or } x = a^2R^k(ea^2e)^{-1}a^2 \text{ and } y = a^{2+k}\}$$

is a congruence. The resulting quotient of $\mathbf{V}$ is an INFB semigroup of the type described in Case 1. Therefore we can assume that not both of l and r are 1.

Now we are ready to compare $\mathbf{V}$ to a semigroup from Ξ_2. Define a map

$$f : \{a, a^2, \ldots a^{1+p}\} \to \mathbf{V}_e$$

by $f(a^i) = ea^ie$. For any $i < r$ and any $s \in \mathbf{V}_e$, $a^{2+i}e \neq a^2s$ and likewise for l, there are at most $(2+r) \times (2+l)$ $\mathcal{H}$-classes $H_{i,j}$. To see that there are exactly $(2+r) \times (2+l)$ $\mathcal{H}$-classes of the form $H_{i,j}$, note that if $a^isa^j = a^{i'}ta^{j'}$ with j and j' less than $2+l$ then $ea^isa^j = ea^{i'}ta^{j'}$, that is there is an element $v \in \mathbf{V}_e$ so that $ea^j = va^{j'}$. Say $j' \neq j$. If both j and j' are greater than 1 then because $a^{2+p} = a^2$ we have $ea^2 = va^{2+|j-j'|}$, contradicting the minimality of l. If one of j and j', say j', is less than 2 then, we have either $ea = ea^j$ or $e = ea^j$. In either case we obtain $ea = ea^k$ for some $k > 1$. But then $ea^2 = ea^{1+k}$. By the arguments above, $1+k$ must equal $2+p$ and therefore $ea = ea^{1+p}$, contradicting the fact that INFB occurs at (a,e). Therefore j must equal j'. Likewise by symmetry if i and i' are less than $2+r$ then $i = i'$.

For $i < 2+r$, $j < 2+l$ and $0 \leq k \leq 1+q$ define a map $\iota' : \mathbf{V} \to \Xi_2(\mathbf{V}_e, L, R, f, q)$ by

$$\iota'(a^i s a^j) = (i, s, j),\ \iota'(a^k) = a^k.$$

It is clear that ι' is a bijection. To show it is an isomorphism, up to symmetry there is only one nontrivial case to check that is not already covered by corresponding arguments for the map ι in Case 1. The case that remains is when $\iota'((a^k)(a^i s a^j)) = \iota'(a^k)\iota'(a^i s a^j)$. If $(k+i) \geq 2$, the left side of this equals

$$\begin{aligned}
&\iota'(a^{2+((k+i-2)mod(r))} R^{[(k+i-2)/r]} s a^j) \\
&= (2 + ((k+i-2)mod(r)), R^{[(k+i-2)/r]} s, j) \\
&= a^k(i, s, j) \ \text{(by Note 3.3.6)} \\
&= \iota'(a^k)\iota'(a^i s a^j)
\end{aligned}$$

as required. If $(k+i) < 2$ we can assume that $k = 1$ and $i = 0$ (since the cases when $k = 0$ are trivial) and the left side becomes

$$\iota'((a)(sa^j)) = (1, s, j) = a(0, s, j) = \iota'(a)\iota'(sa^j)$$

as required. Therefore $\mathbf{V}$ is isomorphic to $\Xi_2[\mathbf{V}_e, L, R, f, q]$. The proof is complete. □

We now have the following Theorem.

THEOREM 3.3.10 *There are infinitely many minimal finitely generated INFB varieties.*

Proof: We will only consider the small INFB semigroups of the first kind. It is well known that if $p > 2$ is a prime number then the dihedral group $\mathbf{D_p}$ given by $\langle a, b : a^p = 1 = b^2, a^{p-1}b = ba \rangle$ is centreless (the proof of this and more general results are popular exercises in many group theory texts; see [3] or [72]). Let $\mathbf{S}$ and $\mathbf{T}$ be two monoids from Ξ_1 with largest subgroup $\mathbf{D_p}$ and $\mathbf{D_q}$ respectively (p and q distinct primes). For each number $n \geq 1$, $\mathbf{D_n}$ has exponent $2n$ so therefore

$\mathbf{S} \models x^2 \approx x^{2+2p}$ and $\mathbf{T} \models x^2 \approx x^{2+2q}$. In this case, any semigroup $\mathbf{U} \in \mathcal{V}(\mathbf{S}) \cap \mathcal{V}(\mathbf{T})$ satisfies $x^2 \approx x^{2+g.c.d.(2p,2q)} \equiv x^{2+2}$. Any subgroup $\mathbf{G}$ of $\mathbf{U}$ must therefore have exponent 2. So $\mathbf{G}$ satisfies $xy \approx xy(yxyx) \approx x(yy)xyx \approx xxyx \approx yx$. That is, $\mathbf{G}$ is abelian, and therefore nilpotent. Therefore $\mathbf{U}$ is INFB if and only if $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{U})$. Since $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$ and $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{T})$, $\mathbf{U}$ is WFB. The result now follows since there are infinitely many prime numbers and consequently infinitely many dihedral groups $\mathbf{D_p}$. □

Note that if INFB occurs at a dividing pair (x, y) in a semigroup $\mathbf{S} \in \Xi_1 \cup \Xi_2$ with maximal subgroup $\mathbf{G}$, then since a is the only non group element in $\mathbf{S}$, x must equal a by Lemma 3.2.1. Every idempotent in $\mathbf{S}$ of the form (i, s, j) for $s \in \mathbf{G}$ and $i + j \geq 1$ can be written in the form $a^i(0, t, 0)a^j$ for some element $t \in \mathbf{G}$. But since $i + j \geq 1$ and the index of $\mathbf{S}$ is 2,

$$\begin{aligned} a^i(0,t,0)a^j aa^i(0,t,0)a^j &= a^i(0,t,0)a^{j+i+1}(0,t,0)a^j \\ &= a^i(0,t,0)a^{j+i+1+d}(0,t,0)a^j \\ &= a^i(0,t,0)a^j a^{d+1}a^i(0,t,0)a^j \end{aligned}$$

and therefore INFB does not occur at $(a, a^i(0, t, 0)a^j)$. So the idempotent y must be one of 1, $(0, e, 0)$ or possibly a^d if $\mathbf{S} \in \Xi_2$. Since INFB occurs at (a, y) it is easily verified that the only possibility for y is $(0, e, 0)$. That is, there is only one dividing pair in $\mathbf{S}$ where INFB occurs. Since INFB occurs for at least two distinct dividing pairs in both $\mathbf{B}_2^1$ and $\mathbf{A}_2^1$, by Lemma 3.1.1 we have proved the following theorem.

THEOREM 3.3.11 *If INFB occurs at only one dividing pair in a finite semigroup* $\mathbf{S}$ *then* $\mathbf{B}_2^1$ *is not contained in the variety* $\mathcal{V}(\mathbf{S})$ *and there is a monoid* $\mathbf{T} \in \Xi_1 \cup \Xi_2$ *such that* $\mathbf{T} \in \mathcal{V}(\mathbf{S})$.

Note also that since the index of every semigroup $\mathbf{S}$ in $\Xi_1 \cup \Xi_2$ is only 2, for any element x, x^2 lies in a subgroup of $\mathbf{S}$. From this it is easily verified that every semigroup from Ξ_1 of period d satisfies $(x^2y)^d \approx (x^3y)^d$ and $(yx^2)^d \approx (yx^3)^d$.

However if $\mathbf{S}$ is a semigroup from Ξ_2 then the numbers l and r are not both 1, and either $a^2(0,e,0)$ and $a^3(0,e,0)$ or $(0,e,0)a^2$ and $(0,e,0)a^3$ must lie in different subgroups of $\mathbf{S}$. In this case one of the identities $(x^2y)^d \approx (x^3y)^d$ and $(yx^2)^d \approx (yx^3)^d$ must fail on $\mathbf{S}$. Thus a minimal finite INFB semigroup in a variety generated by a semigroup from Ξ_1 must be a semigroup from Ξ_1. It is unknown if the same is true for the semigroups in Ξ_2: possibly there are no minimal finite INFB semigroups in Ξ_2 (in which case $\Xi_1 \cup \{\mathbf{B}_2^1\}$ contains all minimal finite INFB semigroups). If however we replace "minimal finite INFB semigroups" with "minimal finite INFB *divisors*" a complete description is possible and indeed, this class contains many small INFB semigroups of the second kind (here a semigroup $\mathbf{S}$ is a divisor of a semigroup $\mathbf{T}$ if $\mathbf{S}$ is a homomorphic image of a subsemigroup of $\mathbf{T}$). To prove this we consider the cases of Ξ_1 and Ξ_2 separately.

If $\mathbf{G}$ is a centreless group and a and b are elements of $\mathbf{G}$ then we will say that a and b are Γ-*separate* in $\mathbf{G}$ if for every proper normal subgroup $\mathbf{N}$, $a\mathbf{N} \equiv b\mathbf{N}$ modulo $\Gamma(\mathbf{G}/\mathbf{N})$. In other words, a and b are distinct modulo $\Gamma(\mathbf{G})$ but in every quotient of $\mathbf{G}$, the cosets containing a and b are equivalent modulo the corresponding upper hypercentre. In a semigroup from Ξ_1 we have that $(1,e,1)a(1,e,1) = (1,h,1)$ and $(1,e,1)a^{d+1}(1,e,1) = (1, g_2^{-1}g_i g_2^{-1}, 1)$. Since INFB occurs at the pair $(a,(1,e,1))$ the group elements h and $g_2^{-1}g_i g_2^{-1}$ must be distinct. This motivates the following definition.

DEFINITION 3.3.12 *Let $\bar{\Xi}_1$ be the subset of Ξ_1 consisting of all monoids of the form $\Xi_1[\mathbf{G}, g, g_1, h]$ so that $(g_1 g)^{-2} g_1^{-1}$ and h are Γ-separate in $\mathbf{G}$ and equivalent modulo $\Gamma(\mathbf{H})$ for every proper subgroup $\mathbf{H}$ of $\mathbf{G}$ containing h and $g_2^{-1}g_i g_2^{-1}$ (for $1 \leq i \leq p$).*

THEOREM 3.3.13 *Every monoid in $\bar{\Xi}_1$ is a minimal INFB divisor for the class of finite semigroups and every minimal INFB divisor for the class of finite semigroups in Ξ_1 is contained in $\bar{\Xi}_1$.*

Proof: Firstly if $\Xi_1[\mathbf{G}, g, g_1, h]$ is not contained in $\bar{\Xi}_1$ then either $(g_1g)^{-2}g_1^{-1}$ and h are not Γ-separate in $\mathbf{G}$ or there is a subgroup $\mathbf{H}$ of $\mathbf{G}$ containing the entries of the sandwich matrix of $\mathcal{M}[\mathbf{G}, 3, 3, P]$ in which $(g_1g)^{-2}g_1^{-1}$ and h are not equivalent modulo $\Gamma(\mathbf{H})$. In the first case, there is a normal subgroup $\mathbf{N}$ of $\mathbf{G}$ so that $(g_1g)^{-2}g_1^{-1}\mathbf{N}$ and $h\mathbf{N}$ are not equivalent modulo $\Gamma(\mathbf{G}/\mathbf{N})$. The quotient $\mathbf{G}/\mathbf{N}$ induces a congruence θ on $\Xi_1[\mathbf{G}, g, g_1, h]$ defined by

$$\{(x, y) : x = y, \text{ or } x = (i, s, j),\ y = (i, t, j), \text{ and } s\mathbf{N} = t\mathbf{N}\}.$$

Since $(g_1g)^{-2}g_1^{-1}\mathbf{N}$ and $h\mathbf{N}$ are not equivalent modulo $\Gamma(\mathbf{G}/\mathbf{N})$, it must be the case that $\Xi_1[\mathbf{G}/\mathbf{N}, g\mathbf{N}, g_1\mathbf{N}, h\mathbf{N}]$ is an INFB divisor of $\Xi_1[\mathbf{G}, g, g_1, h]$.

In the second case since every entry in the sandwich matrix P of $\mathcal{M}(\mathbf{G}, 3, 3, P)$ is an element of $\mathbf{H}$, there is a proper INFB subsemigroup of $\Xi_1[\mathbf{G}, g, g_1, h]$ generated by $\mathbf{H}$, 1 and a. So, again, $\Xi_1[\mathbf{G}, g, g_1, h]$ is not a minimal INFB divisor.

Now assume that $\mathbf{S} = \Xi_1[\mathbf{G}, g, g_1, h]$ is an element of $\bar{\Xi}_1$. Since $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{S})$, Theorem 3.2.11 implies that any congruence on $\mathbf{S}$ whose corresponding quotient $\mathbf{T}$ is INFB must only collapse elements within $\mathcal{H}$-classes. This corresponds to taking a quotient of the group $\mathbf{G}$ in every $\mathcal{H}$-class of $\mathcal{M}[\mathbf{G}, 3, 3, P]$. But since $(g_1g)^{-2}g_1^{-1}$ and h are Γ-separate and $\mathbf{T}$ is INFB, the normal subgroup of $\mathbf{G}$ must be trivial and so $\mathbf{T}$ is isomorphic to $\mathbf{S}$. For similar reasons, the definition of $\bar{\Xi}_1$ and Theorem 3.2.11 imply that there are no proper INFB subsemigroups of $\mathbf{S}$. Therefore $\mathbf{S}$ is a minimal INFB divisor. □

We now investigate semigroups from Ξ_2.

DEFINITION 3.3.14 *Let $\bar{\Xi}_2$ be the subset of Ξ_2 consisting of all monoids of the form $\Xi_2[\mathbf{G}, L, R, f, p]$ so that:*
(i) the elements $f(a^{1+p})$ and $f(a)$ are Γ-separate in $\mathbf{G}$ and equivalent modulo $\Gamma(\mathbf{H})$ for every subgroup $\mathbf{H}$ of $\mathbf{G}$ containing $f(a^i)$ for all $1 \leq i \leq p+1$;
(ii) the numbers r and l (that is $p/(order(R))$ and $p/(order(L))$) are the smallest

choices of i and j respectively with the property that for all $k \leq p$, $f(a^{2+k+i}) = f(a^{2+k})g$ and $f(a^{2+k+j}) = hf(a^{2+k})$ for some elements g and h of $\mathbf{G}$ not dependent on k.

THEOREM 3.3.15 *Every monoid in $\bar{\Xi}_2$ is a minimal INFB divisor for the class of finite semigroups and every minimal INFB divisor in Ξ_2 is contained in $\bar{\Xi}_2$.*

Proof: Let $\mathbf{S} = \Xi_2[\mathbf{G}, L, R, f, p]$ be a semigroup from Ξ_2 for which at least one of the conditions of Definition 3.3.14 is not satisfied. If the first condition is not satisfied then it follows by only trivial modifications of the argument used in the proof of Theorem 3.3.13 that there is a proper INFB divisor of $\mathbf{S}$. So now assume that the first condition holds for $\mathbf{S}$ but the second condition does not. In particular let us assume that there is a smallest number $r' < r$ so that $f(a^{2+k+r'}) = f(a^{2+k})K$ for some $K \in \mathbf{G}$ and for every $k \leq p$. Now $f(a^{2+k})R = f(a^{2+k+r}) = f(a^{2+k+((r)mod(r'))})K^{[r/r']}$ and since $(r)mod(r') < r'$, by the minimality of r' we must have that $(r)mod(r') = 0$ and $K^{[r/r']} = R$. Therefore r' divides r and $RK = KR$. We now show that the equivalence θ given by the symmetric closure of

$$\Delta \vee \{((2+k, Kg, j), (2+((k+r')mod(r)), R^{[(k+r')/r]}g, j)) : \\ 0 \leq k \leq r-1,\ 0 \leq j \leq 1+l,\ g \in \mathbf{G}\},$$

(where Δ denotes the diagonal relation on $\mathbf{S}$) is a congruence so that $\mathbf{S}/\theta$ is INFB. Let $(2+i, Kg, j)$ and $(2+((i+r')mod(r)), R^{[(i+r')/r]}g, j)$ be two θ equivalent elements. Firstly

$$a(2+((i+r')mod(r)), R^{[(i+r')/r]}g, j) \\ = \begin{cases} (2+((1+i+r')mod(r)), R^{[(1+i+r')/r]}g, j), & if\ i < r-1 \\ (2+((r')mod(r)), R^{[r'/r]}Rg, j), & if\ i = r-1, \end{cases}$$

which is equivalent modulo θ to $(2+1+i, Kg, j) = a(2+i, Kg, j)$ if $i < r-1$ and equivalent modulo θ to $(2, KRg, j) = (2, RKg, j) = a(2+i, Kg, j)$ if $i = r+1$.

That $(2+i, Kg, j)a$ and $(2+((i+r')mod(r)), R^{[(i+r')/r]}g, j)a$ are equivalent modulo θ is trivial. Likewise if $k_1, k_2 \leq 1+l$ and $h \in \mathbf{G}$ then $(2+i, Kg, j)(k_1, h, k_2)$ and $(2+((i+r')mod(r)), R^{[(i+r')/r]}g, j)(k_1, h, k_2)$ are also trivially equivalent modulo θ. Now

$$\begin{aligned}
&(k_1, h, k_2)(2+((i+r')mod(r)), R^{[(i+r')/r]}g, j) \\
&= (k_1, hf(a^{k_2+2+((i+r')mod(r))})R^{[(i+r')/r]}g, j) \\
&= (k_1, hf(a^{2+k_2+i+r'})g, j) \\
&= (k_1, hf(a^{2+k_2+i})Kg, j) \\
&= (k_1, h, k_2)(2+i, Kg, j)
\end{aligned}$$

as required. So therefore θ is a congruence on $\mathbf{S}$. By definition θ does not collapse elements within $\mathcal{H}$-classes of $\mathbf{S}$ and so therefore $f(a)$ and $f(a^{1+p})$ are still not equivalent modulo $\Gamma(\mathbf{G})$ and $\mathbf{S}$ is INFB. This means that elements of Ξ_2 that are not elements of $\bar{\Xi}_2$ are not minimal INFB divisors. We now show that elements of Ξ_2 that are not minimal INFB divisors are not elements of $\bar{\Xi}_2$.

Let θ be a congruence on a semigroup $\mathbf{S} = \Xi_2[\mathbf{G}, L, R, f, p]$ from Ξ_2 so that $\mathbf{T} = \mathbf{S}/\theta$ is INFB.

Case 1. $(a^i, a^j) \in \theta$ where $p+1 \geq i, j > 1$ and $i \neq j$.

Since the set $\{a^2, a^3, \ldots, a^{1+q}\}$ is a cyclic subgroup of $\mathbf{S}$ we must have $(a^{k+|i-j|}, a^k) \in \theta$ for all $k \geq 2$. Then $((0, f(a^{k+|i-j|}), 0), (0, f(a^k), 0)) \in \theta$ for all $k \geq 2$. Since for some $k \geq 2$ the elements $f(a^{k+|i-j|})$ and $f(a^k)$ are distinct in $\mathbf{G}$ (by Definition 3.3.5), θ induces a nontrivial congruence on $\mathbf{G}$ and therefore, $f(a)$ and $f(a^{1+p})$ cannot be Γ-separate in $\mathbf{G}$ (since $\mathbf{T}$ is INFB). That is, $\mathbf{S} \notin \bar{\Xi}_2$.

Case 2. $((j, g, k), (j', h, k')) \in \theta$ where (j, g, k) does not equal (j', h, k').

If $j = j'$ and $k = k'$ but $g \neq h$ then clearly the restriction of θ to $\mathbf{G}$ is a nontrivial congruence so $f(a)$ and $f(a^{1+p})$ again cannot be Γ-separate and $\mathbf{S} \notin \bar{\Xi}_2$. If $j \neq j'$ (say, $j < j'$) then both j and j' are greater than 1 (see proof of Case 2 of Theorem 3.3.9). So assume $1 < j < j' \leq r+1$. Now because $(j, g, k) = a^j(0, e, 0)(0, g, k)$ and

$(j', h, k') = a^{j'}(0, e, 0)(0, h, k')$ we have that

$$a^{2+p-j}a^j(0, e, 0)(0, g, k)(0, P_{k,0}^{-1}g^{-1}, 0) = a^2(0, e, 0)$$

and

$$a^{2+p-j}a^{j'}(0, e, 0)(0, h, k)(0, P_{k,0}^{-1}g^{-1}, 0) = a^{2+j'-j}(0, hg^{-1}, 0).$$

By multiplying on the left by $(0, e, 0)a^k$ $(0 \leq k \leq 1 + p)$ we have that

$$((0, e, 0)a^{2+n}(0, e, 0), (0, e, 0)a^{2+k+j'-j}(0, hg^{-1}, 0)) \in \theta$$

and therefore

$$((0, f(a^{2+k}), 0), (0, f(a^{2+k+j'-j})hg^{-1}, 0)) \in \theta$$

for every $k \geq 0$. If $f(a^{2+k}) = f(a^{2+k+j'-j})hg^{-1}$ then condition (iii) implies $\mathbf{S} \notin \bar{\Xi}_2$ (since $j' - j < r$). If $f(a^{2+k}) \neq f(a^{2+k+j'-j})hg^{-1}$ then the congruence θ induces a nontrivial congruence on the group $\mathbf{G}$. Since we have assumed $\mathbf{S}/\theta$ is INFB the group elements $f(a)$ and $f(a^{1+p})$ are not equivalent modulo $\Gamma(\mathbf{G}/\theta)$ and therefore they are also not Γ-separate in $\mathbf{S}$. So $\mathbf{S}$ is not a semigroup from $\bar{\Xi}_2$.

Case 3. $(a^i, (j, s, k)) \in \theta$ for some $i, j, k \leq p + 1$.

By Theorem 3.2.11, $(a, (j, s, k)) \notin \theta$. Say $(a^i, (j, s, k)) \in \theta$ with $i \geq 2$. Since we have that $\{a^2, a^3, \ldots, a^{1+p}\}$ is a subgroup of $\mathbf{S}$, $(a^{i'}, (j, s, k)) \in \theta$ for every $i' \geq 2$. By the arguments used above in Case 2 of the proof of Theorem 3.3.9, we can assume that $j = k = 2$. In accordance with Definition 3.3.5, let l and r be such that p/l and p/r are the orders of L and R respectively. Now at least one of l and r (say r) are greater than 1 and therefore without loss of generality we may assume that $a^2(1, e, 2) = (3, e, 2)$. But since $(a^2, (2, g, 2)) \in \theta$ for some group element $g \in \mathbf{G}$, we must have that $((2, g, 2)(1, e, 2), (3, e, 2)) \in \theta$, that is $((2, gf(a^3), 2), (3, e, 2)) \in \theta$ and therefore $\mathbf{S} \notin \bar{\Xi}_2$ by Case 2 above. □

EXAMPLE 3.3.16 *For any integer $p > 1$ and minimal centreless group divisor $\mathbf{G}$ (say $\mathbf{S}_3$ for example) the semigroup $\Xi_2[\mathbf{G}, e, e, f, p]$ is a minimal INFB divisor if $f(a^i) = e$ for every $i \geq 1$ except when $i = 1 + p$.*

It would have been convenient in Definition 3.3.12 if we had defined elements $a, b \in \mathbf{G}$ to be Γ-separate when in every quotient $\mathbf{H}/\mathbf{N}$ of a subgroup $\mathbf{H}$ of $\mathbf{G}$ containing a and b (that is, in every relevant divisor of $\mathbf{G}$), $a\mathbf{N}$ and $b\mathbf{N}$ are equivalent modulo $\Gamma(\mathbf{H}/\mathbf{N})$. However this choice would make Definition 3.3.14 too complicated since here one needs to account for all the values of $f(a^i)$ and not just $f(a)$ and $f(a^{1+p})$. It is conceivable that there is a group $\mathbf{G}$ and elements $f(a)$, $f(a^2), \ldots, f(a^{1+p})$ satisfying the conditions of Definition 3.3.5 such that in every subgroup $\mathbf{H}$ of $\mathbf{G}$ containing $f(a^i)$ for all $i \leq 1+p$, $f(a)$ and $f(a^{1+p})$ are equivalent modulo $\Gamma(\mathbf{H})$ but also that there is a subgroup $\mathbf{H}'$ containing $f(a)$ and $f(a^{1+p})$ (but not containing at least one element $f(a^i)$) in which $f(a)$ and $f(a^{1+p})$ are not equivalent modulo $\Gamma(\mathbf{H}')$. Under the proposed alternative definition of being Γ-separate, the semigroup $\mathbf{S} = \Xi_2[\mathbf{G}, e, e, f, p]$ might be a minimal INFB divisor even though $f(a)$ and $f(a^{1+p})$ were not Γ-separate.

Combining Lemma 3.1.1, and Theorems 3.3.13 and 3.3.15 we have a description of all minimal finite INFB divisors.

COROLLARY 3.3.17 *The class* $\bar{\Xi}_1 \cup \bar{\Xi}_2 \cup \{\mathbf{B}_2^1, \mathbf{A}_2^1\}$ *is, up to isomorphism, the class of minimal finite INFB divisors.*

Theorem 3.1.2 shows that even though the semigroups from Ξ_1 and Ξ_2 can each be embedded in finite regular semigroups, these semigroups necessarily generate varieties containing $\mathbf{B}_2^1$. Another embedding theorem is that every (finite) semigroup is embeddable in an idempotent generated (finite) semigroup (see [29] for two alternative constructions). As a final observation we show that there are finite INFB idempotent generated semigroups that do not generate varieties containing $\mathbf{B}_2^1$. We use a construction due to T. E. Hall (see [29]). Take an arbitrary semigroup $\mathbf{S}$ from $\Xi_1 \cup \Xi_2$ with period d and therefore satisfying the identity $(xyx^2y)^d \approx (xy)^d$. Construct a Rees matrix semigroup $\mathcal{M}[\mathbf{S}, |\mathbf{S}|, |\mathbf{S}|, P]$ over $\mathbf{S}$ with sandwich matrix satisfying $P_{1,i} = P_{i,1} = 1$ and $\mathbf{S} = \{P_{i,j} : i, j \neq 1\}$. Then $\mathcal{M}[\mathbf{S}, |\mathbf{S}|, |\mathbf{S}|, P]$ is idempotent generated.

PROPOSITION 3.3.18 *The semigroup* $\mathbf{M} = \mathcal{M}[\mathbf{S}, |\mathbf{S}|, |\mathbf{S}|, P]$ *as constructed above is an INFB idempotent generated finite semigroup not generating a variety containing* $\mathbf{B}_2^1$.

Proof: $\mathbf{M}$ is obviously finite and also INFB since $\mathbf{S}$ is embedded in $\mathbf{M}$ (for example $s \mapsto (1, s, 1)$ is an embedding) and $\mathbf{S}$ is INFB. We show that $\mathbf{M} \models (xyx^2y)^d \approx (xy)^d$, where d is the period of $\mathbf{S}$. Since both sides of the identity $(xyx^2y)^d \approx (xy)^d$ start and finish with the same letter, any value these words assume in $\mathbf{M}$ always lies within the same set $M_{i,j} = \{(i, s, j) : s \in \mathbf{S}\}$. Indeed $(xyx^2y)^d$ lies in the same subgroup of $M_{i,j}$ as $(xy)^d$. Now $\mathbf{M}$ has period d and index 2 since $\mathbf{S}$ has this period and index respectively. Therefore $(xyx^2y)^d$ and $(xy)^d$ are both idempotents (note that $d > 2$ since, as noted in the proof of Theorem 3.3.10, a group of period 2 is abelian) and therefore equal. Since $\mathbf{B}_2^1$ does not satisfy $(xyx^2y)^d \approx (xy)^d$ (see proof of Lemma 3.3.3), $\mathbf{B}_2^1 \notin \mathcal{V}(\mathbf{M})$. □

Chapter 4

Finitely generated varieties with uncountably many subvarieties.

By a well known result of Oates and Powell [59], every finite group generates a variety $\mathcal{V}$ with the property that $\mathcal{V}$ and every subvariety of $\mathcal{V}$ can be given by finitely many identities. Such a variety is called *hereditarily finitely based*. Since there are only countably many finite sets of identities, a hereditarily finitely based variety has at most countably many subvarieties (in fact a variety generated by a finite group has only finitely many subvarieties [59]). This situation does not extend to semigroups in general however. In [92] it is shown that the variety generated by $\mathbf{A}_2^1$ has uncountably many subvarieties. Since any variety containing $\mathbf{A}_2^1$ also has uncountably many subvarieties, this example immediately provides a number of finite semigroups, each generating uncountably many subvarieties. However since $\mathbf{A}_2^1$ is INFB so must be every finite semigroup whose variety contains it. This leaves open the question as to the existence of a FB finite semigroup generating a variety with uncountably many subvarieties.

The important semigroup $\mathbf{B}_2^1$ generates a proper subvariety of $\mathcal{V}(\mathbf{A}_2^1)$ so the result of [92] also leaves open the possibility that $\mathcal{V}(\mathbf{B}_2^1)$ generates a variety with only countably many subvarieties. Likewise the small INFB semigroups found in

Chapter 3 could possibly generate such varieties. Note that there does exists a finite semigroup which is not finitely based and generates a variety with only finitely many subvarieties ([75]).

A subsemigroup or homomorphic image of a semigroup $\mathbf{S}$ generates a subvariety of $\mathcal{V}(\mathbf{S})$. In [75] however it is shown that the class of semigroups generating a variety with only finitely many subvarieties is not closed under direct products. Likewise it is natural to ask whether the class of (finite) semigroups generating varieties with countably many subvarieties is closed under direct products.

In Section 4.1 we will use a result proved in [78] to establish a theorem which in turn provides a solution to all of the above questions. It will follow that there exist finite FB semigroups generating varieties with uncountably many subvarieties, that every finite INFB semigroup generates a variety with uncountably many subvarieties and that there are two finite semigroups generating a varieties with finitely many and countably many subvarieties respectively, but whose direct product has uncountably many subvarieties. These examples also show that for any HFB (finite) semigroup $\mathbf{S}_1$ there are HFB finite semigroups $\mathbf{S}_2$ and $\mathbf{S}_3$ such that at least one of $\mathbf{S}_1 \times \mathbf{S}_2$ and $\mathbf{S}_1 \times \mathbf{S}_2 \times \mathbf{S}_3$ is not HFB (note the distinction between the arbitrary HFB semigroup $\mathbf{S}_3$ and the symmetric group $\mathbf{S}_3$ of the previous chapter). For some large classes of semigroups, we will also obtain a complete description of those members generating a variety with uncountably many subvarieties.

In Section 4.2 further examples of varieties with uncountably many subvarieties are found. In particular it is shown that the semigroup $\mathbf{B}_2 \times S(\{a\})$ generates a variety with uncountably many subvarieties.

A universal algebra $\mathbf{S}$ with presentation $\langle A; R\rangle$ (A is a finite alphabet of generators and R is a finite set of relations between words in the alphabet A) within a variety $\mathcal{V}$ is said to have a decidable word problem (relative to $\mathcal{V}$) if there exists an algorithm which determines when two words w_1 and w_2 in the alphabet A are equivalent in $\mathbf{S}$. The variety $\mathcal{V}$ has a decidable word problem if each finitely pre-

sented algebra in $\mathcal{V}$ has a decidable word problem relative to $\mathcal{V}$. A variety $\mathcal{V}$ has a decidable *uniform* word problem if there exists a single algorithm which solves the word problem (relative to $\mathcal{V}$) in all the finitely presented algebras from $\mathcal{V}$. Obviously the decidability of the uniform word problem in $\mathcal{V}$ implies the decidability of the word problem in $\mathcal{V}$. There is an interesting connection between varieties with uncountably many subvarieties and the solvability of the uniform word problem. This connection is examined in Section 4.3 where we use it to construct varieties which have decidable word problem but undecidable uniform word problem.

4.1 A theorem concerning varieties with uncountably many subvarieties

In this section we establish a result concerning monoids generating semigroup varieties whose lattice of subvarieties is uncountable. A number of corollaries follow. In fact the lattice of subvarieties of these varieties contain a continuum of subvarieties in the sense that they contain an uncountable chain with the same ordering as the real numbers. This is so because the lattices involved contain a copy of the lattice of all subsets of the natural numbers. We use an argument from [61] (page 82). If $\mathbb{Q}$ is the set of all rational numbers then for any real number r and with A_r defined as the set $\{q \in \mathbb{Q} : q < r\}$, it is easily seen that $A_{r_1} \subset A_{r_2}$ if and only if $r_1 < r_2$. Thus there is an uncountable chain in the lattice of subsets of $\mathbb{Q}$ and therefore also in the lattice of subsets of the natural numbers. This argument applies to every example of a variety with uncountably many subvarieties in this Chapter.

For each $n > 2$ let L_n be the word

$$y_1x_1x_2x_3x_4y_1y_2x_5y_2y_3x_6y_3 \cdots y_{n-1}x_{n+2}y_{n-1}y_nx_{n+3}x_{n+4}x_{n+5}x_{n+6}y_n.$$

The following result is proved in [78].

LEMMA 4.1.1 *[78] Assume that $L_n \approx w$ is a balanced identity, that w can be deleted to $y_j x_i y_j$ if and only if L_n can be deleted to $y_j x_i y_j$, and that $1 \leq i < j \leq n+6$ implies w deletes to $x_i x_j$. If a substitution θ exists so that $\theta(L_m)$ is a subword of w then $m = n$.*

We now use this lemma to show the following.

THEOREM 4.1.2 *Let Σ be a set of identities closed under deletion. If xyx is an isoterm for Σ then the variety defined by Σ has uncountably many subvarieties.*

Proof: A semigroup $\mathbf{S}$ with a zero element in the signature $\{\cdot, 0\}$ satisfies an identity $u \approx 0$ exactly when it satisfies the semigroup identities $ux \approx yu \approx u$ (x and y are letters not occurring in the word u). For this reason it will be convenient to consider semigroups with zero element to be in the signature $\{\cdot, 0\}$ and satisfying the identities $x0 \approx 0x \approx 0$. This is not essential, but simplifies the arguments to be used. Let $\mathcal{V}$ be a variety defined by a set, Σ, of identities closed under deletion and for which xyx is an isoterm. If M is a subset of the natural numbers, $\mathbb{N}$, then we will take Σ_M to be the set of identities $\{L_n \approx 0 : n \in M\}$. We show that for every subset M of $\mathbb{N}$, $\Sigma \cup \Sigma_M \vdash L_n \approx 0$ if and only if $n \in M$. That is for each pair of subsets P, Q of $\mathbb{N}$, the sets of identities $\Sigma \cup \Sigma_P$ and $\Sigma \cup \Sigma_Q$ define the same subvariety of $\mathcal{V}$ if and only if $P = Q$. Since there are uncountably many subsets of the natural numbers, there are uncountably many subvarieties of $\mathcal{V}$.

Fix some set $M \subseteq \mathbb{N}$ and assume that $\Sigma \cup \Sigma_M \vdash L_m \approx 0$ for some $m \in \mathbb{N}$. By the definition of a derivation of an identity there are words $u_1, \ldots u_n$ with $u_1 \equiv L_m$, $u_n \equiv 0$ and for each $i < n$, u_{i+1} is obtained from u_i by a single application of an identity from $\Sigma \cup \Sigma_M$. The set Σ is closed under deletion and xyx is an isoterm for Σ, so $\Sigma \not\vdash L_m \approx 0$. Therefore we may find a smallest number k such that u_{k+1} is obtained from u_k by an application of an identity from Σ_M. Now since xyx is an isoterm for Σ the words x and xy are also isoterms for Σ. So a letter x_i is linear in u_k if and only if it is linear in L_m. Also every 2-occurring letter y_j in L_m occurs on

either side of a linear letter x_i, that is, L_m deletes to $y_j x_i y_j$. Since xyx is an isoterm, this happens exactly when u_k deletes to $y_j x_i y_j$ and therefore y_j is 2-occurring in u_k also. So $L_m \approx u_k$ satisfies the first two conditions of Lemma 4.1.1. Finally xy is an isoterm for Σ so u_k deletes to $x_i x_j$ if $1 \leq i < j \leq n+6$ and the third condition also holds. Therefore we may apply Lemma 4.1.1 to the identity $L_m \approx u_k$.

Now u_{k+1} is obtained from u_k by an application of an identity of the form L_i for some $i \in M$. So by Lemma 4.1.1, i must equal m and therefore $m \in M$ as required. □

We note also that if xx is an isoterm for a monoid $\mathbf{S}$ and $\mathbf{S}$ satisfies a nontrivial identity of the form $xyx \approx w$ then w must be a nontrivial permutation of the letters in xyx. So w is one of the words xxy or yxx. However it is shown in [70] that either of the identities $xyx \approx xxy$ and $xyx \approx yxx$ define hereditarily finitely based varieties and therefore the variety generated by $\mathbf{S}$ can have only countably many subvarieties. Since xx is an isoterm for a monoid if and only if it has index three or more we have proved the following.

COROLLARY 4.1.3 *A monoid of index three or more generates a variety with uncountably many subvarieties if and only if it does not satisfy $xyx \approx xxy$ or $xyx \approx yxx$ or equivalently if and only if it is not hereditarily finitely based.*

This corollary can also be extracted from the proof of Lemma 7 and Proposition 4 of [78]. These two results of [78] explicitly concern only nonperiodic monoids (monoids which satisfy no identity of the form $x^n \approx x^{n+m}$) and make extensive use the fact (established elsewhere in [78]) that a nonperiodic hereditarily finitely based semigroup necessarily satisfies the implication $e^2 = e \;\&\; f^2 = f \rightarrow ef = efe$ or its dual. While this implication is not always available in the periodic case (for example the variety $\mathcal{N}$ of normal bands does not satisfy this implication and yet by results of Perkins [63] there exist hereditarily finitely based periodic semigroups of arbitrarily large index generating varieties containing $\mathcal{N}$), it has been pointed out to the author by M. Volkov (private communication) that if the condition of being nonperiodic is

replaced by being a monoid of index at least three then this implication is no longer necessary and the corresponding arguments in [78] continue to hold.

The remainder of this section will be concerned with examining the many consequences of Theorem 4.1.2.

The first consequence we investigate is the following.

COROLLARY 4.1.4 *If* $\mathbf{S}$ *is a finite inherently nonfinitely based semigroup then* $\mathcal{V}(\mathbf{S})$ *has uncountably many subvarieties.*

Proof: We use Theorem 1.1.1: $\mathbf{S}$ is a finite INFB semigroup if and only if for every natural number n, Z_n is an isoterm for the identities of $\mathbf{S}$, where $Z_1 \equiv x_1$ and $Z_n \equiv Z_{n-1}x_nZ_{n-1}$. Now Theorem 1.1.2 implies that $\mathbf{S}$ has an INFB subsemigroup, $\mathbf{T}$, with identity. Since $Z_2 \equiv x_1x_2x_1$ is an isoterm for the identities of $\mathbf{T}$, Theorem 4.1.2 applies and therefore $\mathcal{V}(\mathbf{T})$ (and consequently $\mathcal{V}(\mathbf{S})$) has uncountably many subvarieties. □

Finite bases for all monoids of less than 6 elements are established in [14], [15] and [90]. By examining bases of identities described in these papers, it is evident that Theorem 4.1.2 does not apply to any of them: all monoids of order five or less satisfy a nontrivial identity of the form $xyx \approx w(x,y)$ where $w(x,y)$ is a word in the alphabet $\{x,y\}$. A seven element monoid with a finite basis for identities for which Theorem 4.1.2 applies can however be constructed as follows. Recall the definition of the monoid $S(W)$ for a language W (see page 10). It was seen in that chapter that if W is a set of words then $S(W)$ is a monoid for which every word in W is an isoterm. In particular xyx is an isoterm for the monoid $S(\{aba\})$ and therefore by Theorem 4.1.2, the variety generated by $S(\{aba\})$ has uncountably many subvarieties. By Lemma 2.2.8 a finite basis for the identities of $S(\{aba\})$ is the closure under deleting letters of the following set of identities

$$\{xyxzx \approx xxyz,\ xyy \approx yyx,\ xuyvxy \approx xuyvyx,$$
$$xuyxvy \approx xuxyvy,\ xyuxvy \approx yxuxvy\}$$

We have shown the following.

EXAMPLE 4.1.5 *The monoid $S(\{aba\})$ has 7 elements and generates a FB variety with uncountably many subvarieties.*

Note that $\mathbf{B}_2^1$ and $\mathbf{A}_2^1$ each have only 6 elements and generate varieties with uncountably many subvarieties (by Corollary 4.1.4 above or, in the case of $\mathbf{A}_2^1$, by the result in [92]) however they are also NFB.

We now show that the monoid in Example 4.1.5 is quite closely connected to Theorem 4.1.2. Let $\mathbf{S}$ be a semigroup such that the set $Id(\mathbf{S})$ of all identities satisfied by $\mathbf{S}$ satisfy the conditions of Theorem 4.1.2. Since xyx is an isoterm for $\mathbf{S}$, if an identity $u \approx v \in Id(\mathbf{S})$ can be deleted to an identity $u' \approx v'$ where u' is of the form aba (or a subword of this), then $u' \equiv v'$. Therefore $S(\{aba\})$ satisfies every identity in $Id(\mathbf{S})$ and so Theorem 4.1.2 applies to a set Σ of identities only when $S(\{aba\})$ is contained in the variety defined by Σ. We have proved the following theorem.

THEOREM 4.1.6 *A set of identities Σ contains a subset satisfying the conditions of Theorem 4.1.2 if and only if $S(\{aba\})$ is contained in the variety generated by Σ. In this case the variety defined by Σ has uncountably many subvarieties.*

NOTE 4.1.7 *The semigroup obtained from $S(\{aba\})$ by removing the identity element satisfies $x_1x_2x_3x_4 \approx y_1y_2y_3y_4$ and consequently has only finitely many subvarieties.*

If a word w contains a subword of the form xyx then $S(\{w\})$ will generate a variety containing $S(\{aba\})$ and therefore have uncountably many subvarieties. This means that monoids of the form $S(W)$ which generate such varieties are likely to be very common. Indeed we have the following theorem.

THEOREM 4.1.8 *Let W be a non-empty set of words. The following are equivalent:*

(i) the variety $\mathcal{V}(S(W))$ has only countably many subvarieties;
(ii) the variety $\mathcal{V}(S(W))$ has infinitely many but not uncountably many subvarieties;
(iii) $S(W) \models xyx \approx yxx$ or $S(W) \models xyx \approx xxy$;
(iv) either every word in W is, for some $n \geq 1$ and $m > 1$, of one of the forms $a_1a_2 \dots a_n$ and $a_1a_2 \dots a_{n-1}a_n^m$ or every word in W is of one of the forms $a_1a_2 \dots a_n$ and $a_1^m a_2 \dots a_{n-1}a_n$ exclusively (the a_i are distinct letters);
(v) $S(W)$ generates a hereditarily finitely based variety;
(vi) $S(\{aba\}) \notin \mathcal{V}(S(W))$.

Proof: (i)$\Leftrightarrow$(iii). Since both $xyx \approx yxx$ and $xyx \approx xxy$ define hereditarily finitely based varieties, in order to prove the equivalence of the conditions (i) and (iii) we need to show that if $S(W)$ has only countably many subvarieties then it satisfies one of these identities. By Corollary 4.1.3 we need only consider the case when xx is not an isoterm for $S(W)$. If xx is not an isoterm for $S(W)$ then W contains no subwords of the form uu (where u is a word). If it does not contain a subword of the form uvu either (since xx is not an isoterm for $S(W)$, v must be a word distinct from u), then it is a collection of words of the form $a_1a_2 \dots a_n$ (where the a_i are distinct letters) and is easily seen to satisfy $xyx \approx xxy$. If W does contain a subword of the form uvu then xyx is an isoterm for $S(W)$ and so Theorem 4.1.2 implies $S(W)$ does not have countably many subvarieties.

(iii)$\Leftrightarrow$(v). Since $xyx \approx yxx$ and $xyx \approx xxy$ define hereditarily finitely based varieties we need only show that condition (v) implies condition (iii). This follows since a hereditarily finitely based variety necessarily satisfies condition (i) and condition (i) implies condition (iii).

(iii)$\Leftrightarrow$(iv). That condition (iv) implies condition (iii) is easily verified. Now assume that $S(W) \models xyx \approx xxy$ or $xyx \approx yxx$. So W cannot have a subword of the form uvu where $uvu \not\equiv uuv$ or vuu, since then xyx would be an isoterm. Similarly W cannot contain two subwords, one of the form uuv and the other of the form $v'u'u'$ (where $uuv \not\equiv vuu$ or uvu and $v'u'u' \not\equiv u'v'u'$ or $u'u'v'$) since then both

xxy and yxx are isoterms and $S(W)$ would not satisfy condition (iii). Therefore W must satisfy exactly one of the two situations described in (iv).

(i)$\Leftrightarrow$(vi). From the proof of the equivalence of conditions (i) and (iii), the monoid $S(W)$ generates a variety with uncountably many subvarieties if and only if Theorem 4.1.2 applies to the identities of $S(W)$. The equivalence of conditions (i) and (vi) now follows from Theorem 4.1.6.

To complete the proof it remains to show that $S(W)$ at least an infinity of subvarieties. Since W is non-empty, $\mathcal{V}(S(W))$ is a supervariety of $\mathcal{V}(S(\{a\}))$ where a is a single letter. It is trivial to establish that this semigroup variety is given by the identities $\{xy \approx yx, xx \approx xxx\}$. For each n the (n-nilpotent) variety given by

$$\{x_1x_2 \dots x_n \approx y_1y_2 \dots y_n, xx \approx xxx, xy \approx yx\}$$

defines a distinct subvariety of $\mathcal{V}(S(\{a\}))$. The theorem is proved. □

Examples presented in [75] show that the class of semigroups generating varieties with only finitely many subvarieties is not closed under the taking of direct products (or equivalently joins of varieties). Likewise we have the following result:

COROLLARY 4.1.9 *The class of finite semigroups each generating a variety with countably many subvarieties is not closed under direct products. Therefore the class of varieties with countably many subvarieties does not form a sublattice of the class of all varieties.*

Proof: Theorem 4.1.8 shows that $S(\{xyy\})$ and $S(\{xxy\})$ generate varieties with countably many subvarieties. However $S(\{xyy\}) \times S(\{xxy\})$ does not satisfy either of the identities $xxy \approx xyx$ or $yxx \approx xyx$ and the word xx is an isoterm for this monoid. So by Theorem 4.1.2, $S(\{xyy\}) \times S(\{xxy\})$ generates a variety with uncountably many subvarieties. □

In fact the examples used in this corollary show that the join of two hereditarily finitely based varieties generated by finite semigroups can have uncountably many subvarieties. A more striking example is obtained by considering any finite group not

satisfying one of the identities $xxy \approx xyx$ or $yxx \approx xyx$. As mentioned above, the semigroup variety generated by a finite group G has only finitely many subvarieties. If $\mathbf{G}$ does not satisfy $xxy \approx xyx$, say, then the direct product $\mathbf{G} \times S(\{baa\})$ is a monoid of index three not satisfying either of the identities $xxy \approx xyx$ or $yxx \approx xyx$. By Theorem 4.1.2 $\mathbf{G} \times S(\{baa\})$ generates a variety with uncountably many subvarieties (clearly if $\mathbf{G}$ did not satisfy either of the described identities then instead of $S(\{baa\})$ one may take the semigroup $S(\{aa\})$). The smallest group with this property is the symmetric group $\mathbf{S_3}$ with six elements.

In terms of subvarieties however, a quite surprisingly small example is possible. Let $\mathbf{B}$ be the 27 element group with presentation

$$\langle a, b, c : a^3 = b^3 = 1, cb = bc, ac = ca, ab = bac \rangle$$

([8], page 145). This group satisfies neither of the identities $xyx \approx xxy$ or $xyx \approx yxx$ since $aba = baca = baac$, $aab = abac = bacac = baacc$, and baa represent different elements of $\mathbf{B}$. It is also easy to establish that $\mathbf{B}$ can be generated by just the two elements a, b, that it is of exponent 3 and that it is nilpotent of class 2. Indeed it is the only nonabelian group of order dividing 27 that has exponent 3 (see [8]) and is in fact the free Burnside group of exponent 3 on two generators (see [23] for example). Thus every two generated group in the variety of $\mathbf{B}$ (considered either as a semigroup variety or as a group variety) has order dividing 27 and therefore is either isomorphic to $\mathbf{B}$ or is abelian. However, since the identity $xy \approx yx$ involves just two letters, any noncommutative semigroup variety must contain a two generated noncommutative semigroup. Therefore there are no noncommutative proper subvarieties of $\mathcal{V}(\mathbf{B})$. Since the only commutative variety of exponent 3 is that generated by the additive group of integers modulo 3, the lattice of subvarieties of $\mathbf{V}(\mathbf{B})$ is a three element chain (note that every group variety with fewer than three subvarieties is abelian since the atoms in the lattice of semigroup varieties are generated either by a two element semigroup or a cyclic group of prime order; see [18]). We have shown the following.

EXAMPLE 4.1.10 *The lattice of subvarieties of the variety $\mathcal{V}(\mathbf{B})$ is the 3 element chain and the lattice of subvarieties of $\mathcal{V}(S(\{a\}))$ is countable but the lattice of subvarieties of $\mathcal{V}(\mathbf{B} \times S(\{a\}))$ is uncountable.*

The group $\mathbf{B}$ also plays an important role in the examples constructed in [75].

A small, aperiodic (that is, with only trivial subgroups) example of a pair of semigroups generating hereditarily finitely based varieties whose join has uncountably many subvarieties is also possible. Let $\mathbf{L}^1$ be the left zero semigroup with adjoined identity element.

EXAMPLE 4.1.11 *The lattice of subvarieties of the variety $\mathcal{V}(\mathbf{L}^1)$ has five elements and the lattice of subvarieties of $\mathcal{V}(S(\{aab\})$ is countable but the lattice of subvarieties of $\mathcal{V}(\mathbf{L}^1 \times S(\{aab\}))$ is uncountable.*

We now prove this claim. The lattice of band varieties has been completely described in [6], [19] and [22], and it follows that this semigroup generates a variety with only three proper, nontrivial subvarieties (the variety of semilattices, the variety of left zero semigroups and the variety of left normal bands). Since $\mathbf{L}^1$ contains a left zero semigroup it does not satisfy the identity $xyx \approx yxx$. So the direct product $S(\{aab\}) \times \mathbf{L}^1$ is a monoid of index three not satisfying either of the identities $xxy \approx xyx$ or $yxx \approx xyx$ and therefore by Theorem 4.1.2 it generates a variety with uncountably many subvarieties. As seen above, the monoid $S(\{aab\})$ generates a hereditarily finitely based variety.

These examples suggest the following question.

QUESTION 4.1.12 *Do there exist two (finite) semigroups each generating a variety with only finitely many subvarieties whose direct product has uncountably many subvarieties?*

Note that the direct product of the semigroup $\mathbf{L}^1$ above with any finite band generates a variety with still only finitely many subvarieties (in fact from results of [71],

it follows that the direct product of $\mathbf{L}^1$ with any finite group also generates such a variety; see Corollary 3.1.5 below).

Summarising and combining the ideas above we obtain the following theorem.

THEOREM 4.1.13 *(i) For any semigroup* $\mathbf{S}_1$ *(finite or otherwise) there are finite semigroups* $\mathbf{S}_2$ *and* $\mathbf{S}_3$ *generating hereditarily finitely based varieties so that* $\mathbf{S}_1 \times \mathbf{S}_2 \times \mathbf{S}_3$ *generates a variety with uncountably many subvarieties.*
(ii) If $\mathbf{M}$ *is a monoid of index more than two then there is a finite group* $\mathbf{G}$ *generating a hereditarily finitely based variety with only* 3 *subvarieties so that* $\mathbf{M} \times \mathbf{G}$ *generates a variety with uncountably many subvarieties.*
(iii) If $\mathbf{M}$ *is a monoid of index less than or equal to two then either* $\mathbf{M}$ *satisfies both* $xyx \approx xxy$ *and* $xyx \approx yxx$ *or there is a finite semigroup* $\mathbf{S}$ *generating a hereditarily finitely based variety so that* $\mathbf{M} \times \mathbf{S}$ *generates a variety with uncountably many subvarieties.*

Proof: (i) For $\mathbf{S}_2$ and $\mathbf{S}_3$ one can take, for example, the semigroups $\mathbf{L}^1$ and $S(\{aab\})$ or the semigroups $\mathbf{B}$ and $S(\{aa\})$.

(ii) The monoid $S(\{aa\})$ is contained in the variety generated by $\mathbf{M}$ and therefore the claim follows by taking $\mathbf{G}$ to be the group $\mathbf{B}$ above. To obtain a aperiodic example one may replace the group $\mathbf{B}$ in this argument by the direct product of $\mathbf{L}^1$ with its right dual $\mathbf{R}^1$ and obtain a similar result. The semigroup $\mathbf{L}^1 \times \mathbf{R}^1$ generates a band variety with a lattice of subvarieties consisting of 13 elements.

(iii) If $\mathbf{M}$ does not satisfy one of the described identities then one of the semigroups $\mathbf{M} \times S(\{aab\})$ or $\mathbf{M} \times S(\{abb\})$ generates a variety whose identities are closed under deletion, have index three and do not contain either of the identities $xyx \approx xxy$ and $xyx \approx yxx$. By the last part of Theorem 4.1.2, one of these semigroups generates a variety with uncountably many subvarieties. □

In connection with part (iii) of this theorem we note that a monoid of index one satisfying both $xyx \approx xxy$ and $xyx \approx yxx$ is a semilattice of groups (a Clifford

semigroup), each satisfying these identities. This is because by a well known theorem of A. H. Clifford (see [10] or [29]) a semigroup of index one is a semilattice of completely simple semigroups. A completely simple semigroup that is not merely a group cannot not satisfy both the identities $xyx \approx xxy$ and $xyx \approx yxx$ since it contains a divisor isomorphic to either a left or a right zero semigroup. Therefore if $\mathbf{M}$ is a monoid of index one satisfying both of these identities it is a semilattice of groups; obviously every subgroup of $\mathbf{M}$ also satisfies these identities.

A further class for which we can give a complete description of the finite monoids generating varieties with uncountably many subvarieties is the class of orthodox semigroups. The next result follows almost immediately from Corollary 4.1.4, Corollary 3.1.5 of Chapter 3 and existing results.

COROLLARY 4.1.14 *Let* $\mathbf{S}$ *be a finite orthodox monoid with period* p. *The following are equivalent*

(i) $\mathbf{S}$ *has uncountably many subvarieties,*

(ii) $\mathbf{S}$ *has infinitely many subvarieties,*

(iii) $\mathbf{S}$ *is not hereditarily finitely based,*

(iv) $\mathbf{S}$ *is not finitely based,*

(v) $\mathbf{S}$ *is INFB,*

(vi) $\mathbf{B}_2^1 \in \mathcal{V}(\mathbf{S})$,

(vii) $S(\{a\}) \in \mathcal{V}(\mathbf{S})$,

(viii) $\mathbf{S}$ *is not a union of groups.*

(ix) $\mathbf{S} \not\models xyx \approx (xy)^{p+1}x$.

Proof: One of the main results of [71] is that a finite completely regular orthodox semigroup generates an HFB variety with only finitely many subvarieties. This combined with Corollary 3.1.5 and Corollary 4.1.4 implies the equivalence of conditions (i) to (v) above. The equivalence of conditions (iv) to (ix) follows from Corollary 3.1.5 and Theorem 3.1.2. □

This result shows that orthodox monoids always satisfy quite extreme semigroup properties and emphasises the weak connection between the number of subvarieties of a variety and the presence of a finite basis of identities.

NOTE 4.1.15 *If* $\mathbf{S}$ *is a finite orthodox semigroup (not necessarily a monoid) which is not a union of groups then the variety* $\mathcal{V}(\mathbf{S})$ *has infinitely many subvarieties.*

This is because $\mathbf{S}$ contains a non group element a which, since $\mathbf{S}$ is regular, lies in an ideal whose principal factor is an orthodox completely 0-simple semigroup which is not a union of groups. Consider the two semigroups $\mathbf{B_2}$ and $\mathbf{A_2}$. If $\mathbf{C}$ is a completely 0-simple semigroup that is not a union of groups then there is a subsemigroup of a quotient of $\mathbf{C}$ that is isomorphic to either $\mathbf{B_2}$ or $\mathbf{A_2}$. Since $\mathbf{B_2} \in \mathcal{V}(\mathbf{A_2})$ it must be the case that $\mathbf{B_2} \in \mathcal{V}(\mathbf{S})$ (in fact $\mathbf{A_2}$ contains idempotents whose product is not an idempotent and therefore cannot be contained in the variety of $\mathbf{S}$ anyway). A finite basis for the identities of $\mathbf{B_2}$ has been found by A. N. Trahtman (see page 46 of [82]): it is the set

$$\{x^2 \approx x^3,\ x^2y^2 \approx y^2x^2,\ xyx \approx xyxyx\}.$$

Since every identity in this set contains a letter that occurs at least twice on both sides, they are never applicable to any identity of the form $x_1x_2 \dots x_n \approx y_1y_2 \dots y_n$. Thus by adjoining an identity of this form to the above set of identities, a proper subvariety of $\mathcal{V}(\mathbf{B_2})$ is obtained. Since there are infinitely many such identities and each describes a distinct variety it follows that the variety $\mathcal{V}(\mathbf{B_2})$ contains infinitely many subvarieties. Thus a finite orthodox semigroup containing a non group element always generates a variety with infinitely many subvarieties.

We finish this section with two final applications of Theorem 4.1.2. Let S_n be the semigroup variety generated by all semigroups of order n and M_n be the semigroup variety generated by all monoids of order n. Naturally, $\mathsf{M}_n \subseteq \mathsf{S}_n$.

COROLLARY 4.1.16 M_n, *and consequently* S_n, *has uncountably many subvarieties for* $n > 3$. *For* $n < 3$, M_n *and* S_n *have at most countably many subvarieties.*

Proof: If $n \geq 4$, M_n contains the following: the three element monoid $\mathbf{L}^1$ (the two element left zero semigroup with adjoined identity element); its right zero counterpart $\mathbf{R}^1$; and the four element monoid $S(\{aa\})$. Therefore M_n contains the direct product of these. Since xx is an isoterm for $S(\{aa\})$, $\mathbf{L} \not\models xyx \approx yxx$, and $\mathbf{R} \not\models xyx \approx xxy$, Theorem 4.1.2 now applies. Up to isomorphism there are only two, two element monoids (the two element group and the two element semilattice) and these are both commutative. There are five, two element semigroups (the two previously mentioned along with the two element null semigroup and the two element left and right zero semigroups) and it is trivial to verify that these all satisfy the identities $xyzw \approx xzyw$ and $x^2 \approx x^4$. Therefore both M_n and S_n generate hereditarily finitely based varieties and consequently have countably many subvarieties (see [63]). □

The following question remains unanswered

QUESTION 4.1.17 *Do* M_3 *and* S_3 *have uncountably many subvarieties?*

It can be checked that $xyx \approx xyx^7$ is an identity for both of these varieties. For a list of all semigroups of order three the reader is referred to [65].

4.2 Further varieties with uncountably many subvarieties

The proof of Corollary 4.1.4 depends on the fact that every INFB finite semigroup contains an INFB submonoid. If a locally finite INFB semigroup is infinite then this need not be the case. A particularly important example, $\mathbf{Z}_\infty$, is that obtained by taking the Rees quotient of a free semigroup with respect to the ideal consisting of all words that are not subwords of a Zimin word. It is shown in [74] that a locally finite semigroup whose variety $\mathcal{V}$ contains only WFB groups is INFB if and only if $\mathbf{Z}_\infty$ is contained in $\mathcal{V}$. Thus for varieties with only WFB groups, $\mathbf{Z}_\infty$ is the unique minimum INFB variety. It follows from results in [2], [97] and [74] however that $\mathbf{Z}_\infty$

satisfies the identity $xy^2 \approx zy^2$ and so xyx is not an isoterm for any monoid in the variety $\mathcal{V}(\mathbf{Z}_\infty)$. Therefore Theorem 4.1.2 cannot be applied to the semigroup $\mathbf{Z}_\infty$. We now prove the following result which establishes that the variety $\mathcal{V}(\mathbf{Z}_\infty)$ also has uncountably many subvarieties.

THEOREM 4.2.1 *Let Σ be a set of identities. If for every n the Zimin word Z_n is an isoterm for Σ then the variety defined by Σ has uncountably many subvarieties.*

Proof: The proof uses a similar method to that of Theorem 4.1.2. We construct some words L_n so that the identities of $\mathbf{Z}_\infty$ combined with some set $\{L_m \approx 0 : m \in M \subseteq \mathbb{N}\}$ cannot be applied to derive any identity $L_n \approx 0$ if $n \notin M$. In some sense the proof in this case is simpler than that for Theorem 4.1.2 since the words L_n will turn out to be isoterms for $\mathbf{Z}_\infty$ which is not necessarily the case for the corresponding words in Theorem 4.1.2.

Before continuing the proof we introduce a definition and list some properties of Zimin words.

DEFINITION 4.2.2 *(i) If $w \equiv uv$ is a word then $[u|w$ is the word v (that is, we have removed the initial segment u) and $w|v]$ is the word u (that is, we have removed the final segment v);*
(ii) (following the notation of [73]) If u and $v \equiv x_1x_2\ldots x_n$ are words (the x_i's not necessarily distinct) then

$$[u,v]_b^a \equiv u^a x_1 u x_2 u \ldots u x_n u^b;\ a,b \in \{0,1\}.$$

(here, if w is a word then we take w^0 to be the empty word).

Note that there is only a superficial similarity between the words denoted by $[u,x]_1^0$ and by $\lfloor u,t \rfloor$ since the former denotes a word in which the letter x occurs $|u|$ times whereas the latter denotes a word in which a distinct linear letter t_i is placed between every successive pair of letters in u.

A few simple facts concerning Zimin words may help the reader (for convenience we will take Z_0 to be the empty word).

NOTE 4.2.3 *(i)* $Z_1 \equiv x_1$, $Z_2 \equiv x_1x_2x_1$, $Z_3 \equiv x_1x_2x_1x_3x_1x_2x_1$, *etc.*

(ii) Z_n *is* $2^n - 1$ *letters long.*

(iii) If θ *is the substitution defined by* $\theta(x_i) = x_{m+i}$ $(m,\ i \geq 1)$ *then for* $n \geq m$, $Z_{2n} \equiv [\theta(Z_{n-m}), Z_m]_1^1$

(iv) $Z_{2n} \equiv [Z_0|Z_2[Z_2|Z_4[Z_4|Z_6 \ldots [Z_{(2n-2)}|Z_{2n}$.

(v) Every subword of a Zimin word contains a variable that is 1-occurring.

(vi) For every n, *there are no subwords of* Z_n *of the form* uu *(*u *is a word).*

For more information on Zimin words and proofs of some of these facts the reader should consult [2], [97], or [73].

We now define the words which we will be considering. For each $n \in \mathbb{N}$, let L_n be the word denoted by

$$z_1t_1t_2z_1x_1y_1x_1x_2y_2x_2 \ldots x_ny_nx_nz_2t_3t_4z_3.$$

Note that these words are very similar to the words used in Theorem 4.1.2. It is shown in [69] that these words are independent in the sense that for any distinct natural numbers n and m there is no substitution θ so that L_m contains $\theta(L_n)$ as a subword. Thus if the word L_i is an isoterm for $\mathbf{Z}_\infty$ for all numbers $i > 0$ then for any two distinct subsets P and Q of the natural numbers, the sets $Id(\mathbf{Z}_\infty) \cup \{L_n \approx 0 : n \in P\}$ and $Id(\mathbf{Z}_\infty) \cup \{L_n \approx 0 : n \in Q\}$ define distinct varieties. As in Theorem 4.1.2 this shows that $\mathcal{V}(\mathbf{Z}_\infty)$ has uncountably many subvarieties.

In the following table we define a substitution θ of subwords of Zimin words for the letters in L_n so that $\theta(L_n)$ is itself a subword of the Zimin word Z_{2n+4}. We will assume that for any letter x not in the content $c(L_n)$ of L_n, the assignment θ assigns x some letter that is never a subword of any Zimin word.

x	$\theta(x)$
x_i	$x_{2i-1}Z_{2i-2}$
y_i	$x_{2i}Z_{2i-2}$
z_1	x_{m+1}
z_2	x_{m+3}
t_1	$Z_m\|x_1]$
t_2	$x_1x_{m+2}Z_m$
t_3	$Z_mx_{m+1}x_1x_2x_1x_3$
t_4	$[\theta(t_3)\|Z_{m+2}x_{m+4}Z_{m+2}$
	$\equiv [Z_mx_{m+1}x_1x_2x_1x_3\|Z_{m+2}x_{m+4}Z_{m+2}$

Now

$$\theta(x_iy_ix_i) \equiv x_{2i-1}Z_{2i-2}x_{2i}Z_{2i-2}x_{2i-1}Z_{2i-2} \equiv [x_{2i-1}x_{2i}x_{2i-1}, Z_{2i-2}]_1^0 \equiv [Z_{2i-2}|Z_{2i}.$$

Let A_n denote the subword of L_n given by $x_1y_1x_1x_2y_2x_2\ldots x_ny_nx_n$. So we have

$$\begin{aligned}\theta(A_n) \equiv&\theta(x_1y_1x_1)\theta(x_2y_2x_2\ldots x_ny_nx_n)\\ \equiv&[Z_0|Z_2\theta(x_2y_2x_2\ldots x_ny_nx_n)\\ \equiv&[Z_0|Z_2[Z_2|Z_4\theta(x_3y_3x_3\ldots x_ny_nx_n)\\ \equiv&\ldots \equiv [Z_0|Z_2[Z_2|Z_4\ldots[Z_{2n-2}|Z_{2n}\\ \equiv&Z_{2n}, \text{ by Lemma 4.2.3.}\end{aligned}$$

Also

$$\theta(t_1t_2) \equiv Z_mx_{m+2}Z_m, \text{ and } \theta(t_3t_4) \equiv Z_{m+2}x_{m+4}Z_{m+2}.$$

So finally we have that

$$\theta(L_n) \equiv \overbrace{(x_{m+1})}^{\theta(z_1)}\overbrace{(Z_mx_{m+2}Z_m)}^{\theta(t_1t_2)}\overbrace{(x_{m+1})}^{\theta(z_1)}\overbrace{(Z_m)}^{\theta(A_n)}\overbrace{(x_{m+3})}^{\theta(z_2)}\overbrace{(Z_{m+2}x_{m+4}Z_{m+2})}^{\theta(t_3t_4)}\overbrace{(x_{m+3})}^{\theta(z_2)}$$

and this is a subword of the Zimin word Z_{m+4}.

Let ϕ be a substitution and $p \approx q$ be an identity such that $\phi(p)$ is a subword of L_n and $p \approx q$ does not imply any nontrivial identity of the form $Z_k \approx W$ (clearly any identity of $\mathbf{Z}_\infty$ satisfies this second property). Whenever $\phi(p)$ is a subword of Z_k, we must have $\phi(p) \equiv \phi(q)$. Let $L_n \equiv u\phi(p)v$ and $L'_n \equiv u\phi(q)v$. We have that $(\theta \circ \phi)(p)$ is a subword of Z_{m+4} because $\theta(u\phi(p)v) = \theta(u)(\theta \circ \phi)(p)\theta(v)$. Therefore $(\theta \circ \phi)(p) \equiv (\theta \circ \phi)(q)$ and so $\theta(u\phi(q)v) \equiv \theta(u\phi(p)v)$. The proof will therefore be complete if we can show that $\theta(u\phi(q)v) \equiv \theta(u\phi(p)v)$ implies that $u\phi(p)v \equiv u\phi(q)v$ since this in turn shows that $\phi(p) \equiv \phi(q)$.

Now x_{m+4} occurs just once in $\theta(L_n)$ so the same is true in $\theta(L'_n)$ (since these are identical). The only letter assigned a word containing x_{m+4} by θ is t_4. Furthermore, to the right of $\theta(t_4)$ in $\theta(L_n)$ (and therefore in $\theta(L'_n)$) there is just one letter, the letter x_{m+3}. So in L'_n to the right of t_4 we must have just one letter and that letter must be assigned the letter x_{m+3} by θ. The only letter assigned x_{m+3} by θ is z_2 and therefore t_4z_2 is a final segment of L'_n.

Now to the left of $\theta(t_4)$ in $\theta(L_n)$ and $\theta(L'_n)$ we have the letter x_3. Thus the letter to the left of t_4 in L'_n must be assigned a word ending in x_3 and the only letter for which this is true is the letter t_3. So $t_3t_4z_2$ is a final segment of L'_n. To the left of $\theta(t_3)$ in $\theta(L_n)$ and $\theta(L'_n)$ we have the letter x_{m+3}. Thus the letter to the left of t_3 in L'_n must be assigned a word ending in x_{m+3} and the only letter for which this is true is the letter z_2. So $z_2t_3t_4z_2$ is a final segment of L'_n.

To the left of $\theta(z_2t_3t_4z_2)$ in $\theta(L_n)$, the letter x_{m+2} occurs just once. There is only one letter assigned by θ a word containing x_{m+2} whose length is less than or equal to $\theta(L_n)|\theta(z_2t_3t_4z_2)]$ and that is t_2. Similar arguments to the above now show that an initial segment of $\theta(L'_n)$ is $z_1t_1t_2z_1$ so therefore only the central portion, $\theta(A_n)$, remains to be examined.

Now the first letter of $\theta(A_n)$ is x_1 since $\theta(A_n) \equiv Z_m$. So the letter to the right of the second occurrence of z_1 in T'_n must be assigned by θ a word beginning in x_1. There are five possibilities: x_1, t_1, t_2, t_3 and t_4 however all but the first two of these

are assigned words longer than the whole of $\theta(A_n)$ and so can be eliminated. Now $\theta(t_3) \equiv Z_m|x_1] \equiv \theta(A_n)|x_1]$ which leaves only the letter x_1 remaining in $\theta(A_n)$. The only letter assigned x_1 by θ is the letter x_1 itself. So then $L'_n \equiv z_1t_1t_2z_1t_1x_1z_2t_3t_4z_2$. This is certainly not possible since we can find a new substitution θ' defined by $\theta'(x) \equiv \theta(x)$ if $x \in \{z_1, t_1, t_2, x_1, z_2, t_3, t_4\} = c(L'_n)$ and t otherwise (for some letter t not of the form x_i). Then $\theta'(L'_n) \equiv \theta(L'_n)$ which is a subword of Z_{m+4} but $\theta'(L_n)$ contains the new letter t so is not a subword of Z_{m+4}, contradicting the choice of $p \approx q$.

So therefore the first letter after the second occurrence of z_1 in T'_n is x_1. The next letter in $\theta(A_n)$ is x_2 and the only letter assigned a word starting with x_2 by θ is the letter y_1. Following this in $\theta(A_n)$ we have the letter x_1. This time there is only one letter assigned a sufficiently short word starting with x_1 and that letter is x_1 itself. To the right of this, every new portion of A_n of the form $x_iy_ix_i$ begins with a letter which completely determines a corresponding letter x_i or y_i and thus completely determines the fact that the central portion of L'_n is A_n also. Thus Theorem 4.2.1 is proved since we have shown $L_n \equiv L'_n$. □

By results of [73], a variety $\mathcal{V}$ contains an infinite, finitely generated nil-semigroup (a semigroup satisfying $x^n \approx 0$ for some n) only if $\mathbf{Z}_\infty$ is contained in $\mathcal{V}$. Thus we have the following corollary.

COROLLARY 4.2.4 *Any variety $\mathcal{V}$ containing an infinite, finitely generated nil-semigroup has uncountably many subvarieties.*

Theorems 4.2.1 and 4.1.6 show that a semigroup variety containing $\mathbf{Z}_\infty$ or $S(\{aba\})$ respectively has uncountably many subvarieties. We now find a different example of this kind.

THEOREM 4.2.5 *If $\mathcal{V}$ is a variety containing the semigroups $\mathbf{B_2}$ and $S(\{a\})$ then $\mathcal{V}$ has uncountably many subvarieties.*

Proof: Let $\mathbf{S}$ be the semigroup $\mathbf{B_2} \times S(\{a\})$. Since $\mathbf{B_2}$ and $S(\{a\})$ are (up to isomorphism) subsemigroups of $\mathbf{S}$, a variety $\mathcal{V}$ contains $\mathbf{S}$ if and only if it contains both $\mathbf{B_2}$ and $S(\{a\})$ and the semigroup $\mathbf{S}$ satisfies an identity $p \approx q$ exactly when both the subsemigroups $\mathbf{B_2}$ and $S(\{a\})$ satisfy $p \approx q$. It is easily verified that the semigroup $S(\{a\}) \models p \approx q$ if and only if $c(p) = c(q)$ and $occ(x,p) = 1 \Leftrightarrow occ(x,q) = 1$. As noted above, a basis for the identities of $\mathbf{B_2}$ is the set $\{x^2y^2 \approx y^2x^2, xyx \approx xyxyx, x^2 \approx x^3\}$. It is clear that $\mathbf{B_2} \models p \approx q$ implies $c(p) \approx c(q)$ and therefore $\mathbf{S} \models p \approx q$ if and only if both $\mathbf{B_2} \models p \approx q$ and there is no letter t that is linear on one side of $p \approx q$ but nonlinear on the other. We will show via the following lemmas that for every odd number $n > 0$ the word

$$L_n \equiv (z_1t_1t_2z_1)(x_1y_1x_1)(x_2y_2x_2)\ldots(x_ny_nx_n)(z_2t_3t_4z_2)$$

is an isoterm for $\mathbf{S}$ (the condition of being odd here merely serves to reduce in what follows the number of cases necessary to consider). These words were used in the proof of the previous theorem, and thus if the word L_i is an isoterm for $\mathbf{S}$ for all odd numbers $i > 0$. then for any two distinct subsets P and Q of the odd natural numbers, the sets $Id(\mathbf{S}) \cup \{L_n \approx 0 : n \in P\}$ and $Id(\mathbf{S}) \cup \{L_n \approx 0 : n \in Q\}$ define distinct varieties. This shows that $\mathcal{V}(\mathbf{S})$ has uncountably many subvarieties.

It will be convenient to consider the semigroup $\mathbf{B_2}$ as the semigroup on the set $\{a, b, ab, ba, 0\}$ with presentation $\langle a, b : aba = a, bab = b, aa = bb = 0\rangle$. It is clear that any word in the alphabet $\{a, b\}$ that starts with the letter a represents in $\mathbf{B_2}$ one of the words a, ab or 0 and likewise words starting with b represent one of the words b, ba or 0. The following two lemmas establish the structure of possible words r for which $\mathbf{B_2} \models L_n \approx r$.

LEMMA 4.2.6 *If* $\mathbf{B_2} \models L_n \approx r$ *then* r *begins with the letter* z_1 *and ends with the letter* z_2.

Proof: For every number i less than n assign a to the letters x_{2i-1}, y_{2i}, t_1 and t_3, b to the letters x_{2i}, y_{2i-1} and z_i, and ba to t_2 and t_4. Call this assignment θ_1. Un-

der θ_1, L_n takes the value $[(b)(a)(ba)(b)](aba)(bab)\dots(aba)[(b)(a)(ba)(b)] = b$. Since $\mathbf{B_2} \models L_n \approx r$, the word r must also be assigned the value b under θ_1. This shows that r cannot start with any of the letters x_{2i-1}, y_{2i}, t_1 or t_3. Let θ_2 be the same as θ_1 except with ab assigned to z_1, and b assigned to t_2. This gives L_n the value ab and shows that r cannot start with any of the letters x_{2i}, y_{2i-1}, z_2, t_2 and t_4. Thus r starts with the letter z_1. By the symmetry of the word L_n and of the semigroup $\mathbf{B_2}$ there are dual assignments to the above that show that r must finish with the letter z_2. □

LEMMA 4.2.7 *If* $\mathbf{B_2} \models L_n \approx r$ *and* u *is a two letter subword of* r *then either* u *is a two letter subword of* L_n *or* u *is contained in the set* $\{y_1t_1,\ y_iy_{i-1},\ t_4y_n : 0 < i \leq n\}$.

Proof: Since bb and aa equal zero in the semigroup $\mathbf{B_2}$, the assignments θ_1, θ_2 and their duals above show that the only possible two letter subwords involving letters of the form x_i and y_j are $x_{2i}y_{2j}$ or its reverse, $x_{2i-1}y_{2j-1}$ or its reverse, $x_{2i}x_{2j-1}$ or its reverse, and $y_{2i-1}y_{2j}$ or its reverse. Assume that r contains the subword of the form $x_{2i}y_{2j}$ or its reverse. Say $i \leq j$ and define an assignment ϕ_{2i} as follows. Assign a to all letters $x_{2i'}$ and $y_{2i'-1}$ with $i' \leq i$ and b to all letters $x_{2i'-1}$ and $y_{2i'}$ with $i' \leq i$. Assign a to all letters $x_{2j'-1}$ and $y_{2j'}$ for $j' > i+1$ and b to all letters $x_{2j'}$ and $y_{2j'-1}$ for $j' > i+1$. Assign ba to x_{2i+1} and y_{2i+1}. Since $2i$ is even and n is odd, ϕ_{2i} assigns the word

$$(x_1y_1x_1)(x_2y_2x_2)\dots(x_{2i}y_{2i}x_{2i})[x_{2i+1}y_{2i+1}x_{2i+1}](x_{2i+2}y_{2i+2}x_{2i+2})\dots(x_ny_nx_n)$$

the value

$$(bab)(aba)\dots(aba)[(ba)(ba)(ba)](bab)\dots(aba) = ba.$$

To complete the definition of ϕ_{2i}, let ϕ_{2i} assign ba to z_1 and z_2, b to t_1 and t_3 and a to t_2 and t_4. An analogous assignment for odd numbers $2i-1$ exists and we will denote this by ϕ_{2i-1}. Now ϕ_{2i} gives L_n the value ba on the semigroup $\mathbf{B_2}$. However

it also assigns any word $x_{2i}y_{2j'}$ (or reverse) the value $aa = 0$ if $i < j'$. Since it has been assumed that r contains the subword $x_{2i}y_{2j}$ for $i \leq j$ and $\mathbf{B_2} \models L_n \approx r$ it must be that $i = j$. In the case when $j \leq i$ the same arguments using the substitution ϕ_{2j} instead of ϕ_{2i} again show that $i = j$. These assignments also show that if $x_{2i}x_{2j+1}$ is a subword of r then $j = i$ and that if $y_{2j+1}y_{2i}$ is a subword of r then $j = i$ (note however that there are no such subwords in L_n). Similarly, using ϕ_{2i-1} one can show that if $x_{2i-1}y_{2j-1}$ (or its reverse), $x_{2i-1}x_{2j}$, and $y_{2i}y_{2j-1}$ are subwords of r then $i = j$.

Thus the only possible two letter subwords of r in the alphabet $\{x_i, y_j : 0 < i, j \leq n\}$ are those already occurring in L_n and subwords of the form y_iy_{i-1}. The arguments above are easily extended to the two letter subwords of r containing any of the letters x_i, y_i, z_i or t_i. It is routine to verify in this case that the only possible two letter subwords of r that do not already occur in L_n are those found above and the words y_1t_1 and t_4y_n. The lemma is proved. □

Recall that we are assuming that $\mathbf{B_2} \models L_n \approx r$ and that $\mathbf{S} = \mathbf{B_2} \times S(\{a\})$. Denote the set of all possible two letter subwords of r by $\mathbf{R}$ (note that not all of these subwords need occur in any particular choice of the word r). We now complete the proof of Theorem 4.2.5 by showing that if $\mathbf{S} \models L_n \approx r$ then $L_n \equiv r$.

We associate with the word r a sequence of consecutive edges, or a pathway, in a directed graph $G(r)$ with vertex set $V(G(r)) = c(r) \cup \{0\}$ and edge set $E(G(r)) = \{(u, v) : uv \in \mathbf{R}\} \cup \{(0, z_1), (z_2, 0)\}$ (no duplicate edges are allowed). This graph is shown in Figure 4.1 (here the dotted lines represent edges corresponding to the two letter subwords contained in $\mathbf{R}$ but not occurring in the word L_n). The first edge in the pathway corresponding to r is the edge $(0, z_1)$ and successive edges correspond to successive two letter subwords in r. That is, the i^{th} edge in this pathway corresponds to the $(i-1)^{th}$ two letter subword to occur in r. Finally, the last edge in the pathway is the edge $(z_2, 0)$. Naturally for some choices of r the corresponding pathway does not contain every edge. For example, the word L_n (which is a possible choice for r since $\mathbf{S} \models L_n \approx L_n$ trivially) corresponds to the (unique) pathway passing every

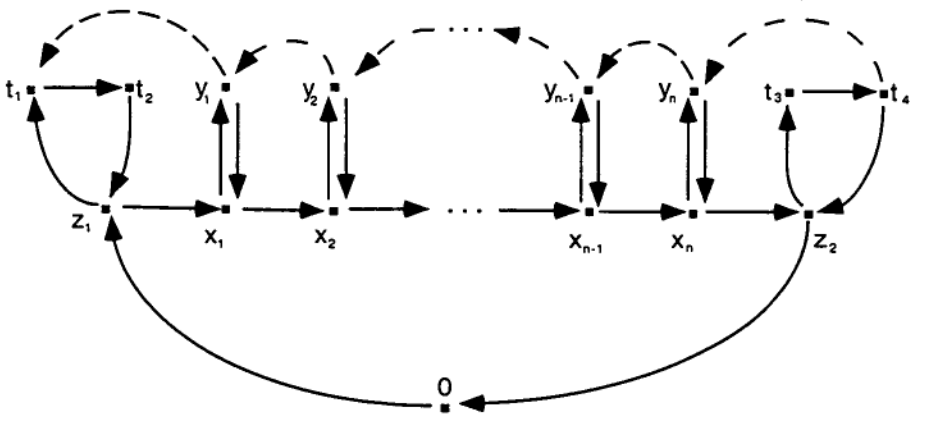


Figure 4.1: The directed graph constructed for the word r.

non dotted edge exactly once. If the semigroup $\mathbf{S} \models L_n \approx r$ then all linear letters in L_n are linear in r also. Therefore for every linear letter, say t, the pathway corresponding to r contains only one edge leaving the vertex t and one entering the vertex t. We will assume that this pathway contains a dotted edge (that is, r contains a two letter subword not contained in L_n) and show that a contradiction arises.

Assume that the edge (y_i, y_{i-1}) is contained in the pathway corresponding to r and that i is the largest number with this property. Thus either the edge immediately preceding (y_i, y_{i-1}) is (x_i, y_i) or $i = n$ and the edge immediately preceding (y_i, y_{i-1}) is (t_4, y_n). Let j be the smallest number for which (y_{j+1}, y_j) is an edge succeeding (y_i, y_{i-1}) in the pathway. Therefore either the edge immediately following (y_{j+1}, y_j) is (y_j, x_j) or $j = 1$ and the edge immediately following (y_{j+1}, y_j) is (y_1, t_1). For the sake of simplicity we will only consider the cases when i does not equal n and j does not equal 1. The remaining cases follow in the same manner essentially by using z_1 and z_2 instead of x_j and x_i respectively (aside from simple arguments regarding t_2 and t_3). So r contains the subword $x_i y_i y_{i-1} y_{i-2} \ldots y_{j+1} y_j x_j$. The only edges pointing left in the graph are of the form (y_k, y_{k-1}), (y_1, t_1) and (t_4, y_n). Thus if an edge of the form (y_k, x_k) is contained in the pathway corresponding to r then, since y_k is

linear in r, every edge to follow can never finish at the vertex x_k. Therefore r must be of the form

$$[A]x_jx_{j+1}\dots x_{i-1}x_iy_iy_{i-1}\dots y_{j+1}y_jx_jx_{j+1}\dots x_{i-1}x_ix_{i+1}[B]$$

where A does not contain x_k or y_k for any $k \geq j$ or the letters z_2, t_3 and t_4 and B does not contain $x_{k'}$ or $y_{k'}$ for any $k' < i+1$ or the letters z_1, t_1 and t_2. Assign ab to all letters in r up to (but not including) the first occurrence of the letter x_i, assign a to x_i, ba to y_i and b to y_{i-1}. Assign ab to the letters y_k for $i-1 < k \leq j$ and ba to all other letters. Clearly (since ab and ba are idempotent in $\mathbf{B_2}$) these rules assign A the value ab and B the value ba. Thus r is assigned the value

$$[ab](ab)(ab)\dots(ab)(a)(ba)(b)(ab)(ab)\dots(ab)(ab)(ab)(ab)\dots(ab)(a)(ba)[ba] = a.$$

However L_n contains the subword $x_{i-1}y_{i-1}$ which takes the value $abb = 0$ under this assignment. Thus we have reached a contradiction.

So the pathway corresponding to r does not pass along any of the dotted edges but does pass through every vertex. Since the vertices $t_1, \dots, t_4$ and $y_1, \dots, y_n$ can be passed only once, it is easily verified that the pathway corresponding to r must be identical to that of L_n. Thus $r \equiv L_n$ as required. □

It is a routine exercise to verify that both $\mathbf{B_2}$ and $S(\{a\})$ satisfy the identity $xyxzx \approx xzxyx$ but $S(\{aba\})$ does not and therefore $S(\{aba\}) \notin \mathcal{V}(\mathbf{B_2} \times S(\{a\}))$. Similarly $\mathbf{B_2} \times S(\{a\}) \notin \mathcal{V}(S(\{aba\}))$ since $S(\{aba\}) \models xyxy \approx yxyx$ but $\mathbf{B_2} \times S(\{a\}) \not\models xyxy \approx yxyx$. It is also evident that the direct product of $\mathbf{B_2}$ with $S(\{a\})$ is not INFB (see Theorem 1.1.2 for example) and therefore $\mathbf{Z}_\infty \notin \mathcal{V}(\mathbf{B_2} \times S(\{a\}))$.

Proposition 3 of [78] shows that if $\mathcal{V}$ is a nonperiodic variety then $\mathcal{V}$ is HFB only if the regular elements of every semigroup $\mathbf{S}$ in $\mathcal{V}$ lie in subgroups of $\mathbf{S}$. To prove this result it is shown that a semigroup $\mathbf{T}$ containing a nongroup, regular element generates a variety containing either $\mathbf{B_2}$ or the bicyclic semigroup with presentation $\langle p, q : pq = 1\rangle$. It is then shown that if a variety $\mathcal{V}$ contains either $\mathbf{B_2}$ or the bicyclic semigroup and is non periodic then $\mathcal{V}$ is not HFB. In fact the condition of

nonperiodicity here serves only to ensure that certain identities are balanced. The identities in question will also be balanced if (as in the comments after Corollary 4.1.3) the condition of being nonperiodic is replaced by the condition of containing a monoid of index more than three. Thus the regular elements of a semigroup in a hereditarily finitely based variety containing a monoid of index at least four are all group elements. We obtain the following improvement on these results.

COROLLARY 4.2.8 *If $\mathcal{V}$ is an hereditarily finitely based variety that contains a monoid of index greater than one (that is, a monoid that is not completely regular) then the regular elements of any semigroup $\mathbf{S}$ in $\mathcal{V}$ lie in subgroups of $\mathbf{S}$. On the other hand if a variety $\mathcal{V}$ contains a semigroup with a nongroup, regular element and $\mathcal{V}$ also contains a monoid of index greater than one then $\mathcal{V}$ has uncountably many subvarieties.*

Proof: The arguments used to prove Proposition 3 of [78] (described above) show that if a variety $\mathcal{V}$ contains a semigroup with a nongroup, regular element then $\mathcal{V}$ contains either $\mathbf{B}_2^1$ or the bicyclic semigroup. In the first case, if $\mathcal{V}$ also contains a monoid of index greater than one, Theorem 4.2.5 implies that $\mathcal{V}$ has uncountably many subvarieties and so cannot be HFB. Now the bicyclic semigroup is a monoid with identity element 1 and is nonperiodic (since, for example, $p^n = p^m$ if and only if $n = m$). Therefore by Corollary 4.1.3 it generates a HFB variety if and only if xyx is an isoterm for its identities, that is, if and only if it does generate a variety with uncountably many subvarieties. However it is known that the bicyclic semigroup is NFB [86] and therefore not HFB. The theorem now follows. □

This theorem provides an example of a seven element, not INFB semigroup whose identities are not closed under deletion. We may think of $\mathbf{B}_2$ and $S(\{c\})$ as sharing a single common element, the zero element (here we use the letter c in the semigroup $S(\{c\})$ to avoid confusion between elements of $S(\{c\})$ and elements of $\mathbf{B}_2$) and define a semigroup multiplication on the set $\mathbf{B}_2 \cup (S(\{c\}))$ to coincide with that on the subsemigroups $\mathbf{B}_2$ and $S(\{c\})$ and to equal zero elsewhere (this

construction is called the *zero direct join* of $\mathbf{B_2}$ and $S(\{c\})$; see also page 31).

EXAMPLE 4.2.9 *The seven element semigroup* $\mathbf{B_2} \cup (S(\{c\}))$ *with the described multiplication generates a variety with uncountably many subvarieties.*

It is trivial to verify that this semigroup has seven elements and generates a variety satisfying the conditions of Theorem 4.2.5 (it generates the same variety as $\mathbf{B_2} \times S(\{c\})$). This semigroup also satisfies $xyxzx \approx xzxyx$ and so is not INFB by results of [73] since this identity implies that the word Z_3 is not an isoterm (see proof of Corollary 4.1.4 above). The identities of this semigroup are not closed under deletion since $xyxzx \approx xzxyx$ deletes to $yz \approx zy$ and $\mathbf{B_2}$ is not commutative. Indeed since the identity $xy \approx yx$ defines a hereditarily finitely based variety (see [63]), this argument shows that any subvariety of $\mathcal{V}(\mathbf{B_2} \times S(\{c\}))$ whose identities are closed under deletion has only countably many subvarieties.

A more extreme example is the semigroup $\mathbf{Z}_\infty$ above.

EXAMPLE 4.2.10 *The semigroup variety generated by* $\mathbf{Z}_\infty$ *has uncountably many subvarieties but contains no nontrivial monoids.*

Proof: As was noted, this semigroup can be shown to satisfy the identity $x^2y \approx x^2z$ and so it follows that any monoid in this variety must satisfy $y \approx z$ and therefore must be trivial. □

A similar example is that found by J. Jezek in [36].

EXAMPLE 4.2.11 *[36] The variety* $\mathcal{V}'$ *defined by* $x^2y \approx yx^2 \approx x^2$ *has uncountably many subvarieties but contains no nontrivial monoids. This variety is the variety of all semigroups where the square of any element is the zero element.*

Proof: That $\mathcal{V}'$ has uncountably many subvarieties is the main result of [36]. Now if 1 is the identity element of a monoid then $1^2 = 1$ and it follows that if s is an element of a monoid $\mathbf{S}$ from $\mathcal{V}'$, then $s = s1^2 = 1^2 = 1$. That is, all monoids in $\mathcal{V}'$ are trivial. □

The variety $\mathcal{V}'$ however is not generated by a finite semigroup, indeed it contains the well known three generated infinite semigroup constructed by Morse and Hedlund [55] and so is not even locally finite. That it has uncountably many subvarieties therefore also follows from Corollary 4.2.4 above.

Note also that Theorem 4.2.5 shows that the direct product $\mathbf{B_2}$ with any monoid of index greater than zero generates a variety with uncountably many subvarieties. If $\mathbf{B_2}$ generates a hereditarily finitely based variety then we would obtain an improvement of Theorem 4.1.13. As an inverse semigroup in the signature $\{\cdot, ^{-1}\}$, $\mathbf{B_2}$ does generate such a variety [40]. This motivates the following question.

QUESTION 4.2.12 *Does* $\mathbf{B_2}$ *generate a hereditarily finitely based semigroup variety?*

The word xyx is an isoterm for all examples found above. On the other hand a recent result of J. Kadourek [38] shows that the semigroup variety defined by the identity $x^2y \approx xy$ has uncountably many subvarieties. Clearly xyx is not an isoterm for this variety. We now present a second example with this property which permits a proof along similar lines to others in this thesis. However it is not known whether the example in [38] or the example below can be modified to imply the existence of *finite* semigroups whose varieties have uncountably many subvarieties. For instance, the variety defined by $x^2y \approx xy$ contains the variety of all bands and therefore by a result from [79], cannot be generated by any finite semigroup.

For every $k > 0$ let $\mathcal{V}_k$ be the variety defined by $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx\}$. Note that while $x^2y \approx xy \vdash xyx \approx xy^{k+1}x$, the variety defined by $\{x^2y \approx xy, xyxy \approx yxyx\}$ has only countably many subvarieties since these identities imply $xyx \approx xyxyx \approx yxyxx \approx yxx$.

EXAMPLE 4.2.13 *For every* $k > 0$, $\mathcal{V}_k$ *has uncountably many subvarieties.*

Proof: For every $n > 0$ let L_n be the word

$$x_1x_2x_1y_1^2y_2^2 \ldots y_n^2x_3x_4x_3.$$

and R_n be the word

$$x_1x_2x_1y_n^2y_{n-1}^2\dots y_1^2x_3x_4x_3.$$

Fix a subset M of the natural numbers $\mathbb{N}$ and let n be any element of $\mathbb{N}$. We will show that $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx, L_i \approx R_i : i \in M\} \vdash L_n \approx R_n$ only if $n \in M$. As in previous proofs, this implies that the variety $\mathcal{V}_k$ has uncountably many subvarieties.

Let the set $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx, L_i \approx R_i : i \in M\}$ be denoted by Σ_M and assume that $\Sigma_M \vdash L_n \approx R_n$. By the definition of a derivation we can select a number m and pairwise distinct words $u_1, u_2, \dots, u_m$ with $u_1 \equiv L_n$, $u_m \equiv R_n$ such that for each $i \leq m$, there is a substitution θ_i and an identity $p_i \approx q_i \in \Sigma_M$ such that u_{i+1} is obtained from u_i by replacing a subword of the form $\theta_i(p_i)$ in u_i with the subword $\theta_i(q_i)$. Let j be the largest number so that $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx\} \vdash u_1 \approx u_j$. There are only two subwords of L_n of the form xyx and none of the form $xyxy$. Since $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx\} \vdash L_n \approx u_j$ it is easily established by induction on j that for some integers $p, q \geq 0$,

$$u_j \equiv x_1x_2^{pk+1}x_1y_1^2y_2^2\dots y_n^2x_3x_4^{qk+1}x_3.$$

Because this word is not R_n it follows that $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx\} \not\vdash L_n \approx R_n$ and so there exists a number $h \in M$ and a substitution θ such that $u_j \equiv r\theta(L_h)s$ and $u_{j+1} \equiv r\theta(R_h)s$. The first letter of L_h is x_1. Since x_1 is 2-occurring in L_h and $x_1x_2x_1$ is a subword of L_h, there must be a subword of u_j of the form uvu for some words u and v. By inspection, the pair (u, v) is one of the following: (x_1, x_2^{pk+1}), $(x_2^{e_1}, x_2^{f_1})$, (x_3, x_4^{qk+1}), $(x_4^{e_2}, x_4^{f_2})$ (where e_i and f_i are natural numbers satisfying $e_1 + f_1 \leq pk+1$ and $e_2 + f_2 \leq qk + 1$). The second last of these is obviously impossible since then uvu would be a final segment of u_j but uvu must be followed in u_j by $\theta(y_1)$ since this follows $x_1x_2x_1$ in L_h. The last of the possibilities is also impossible since the only letter that occurs twice to the right of x_4 in u_j is x_4 itself. This enforces $\theta(x) = x_4^i$ for every letter $x \in c(L_h)$ (for some i depending on x) and therefore $\theta(L_h) \equiv \theta(R_h)$.

In this case $u_j \equiv u_{j+1}$, contradicting both the choice of j as the largest such that $\{xyx \approx xy^{k+1}x, xyxy \approx yxyx\} \vdash L_n \approx u_j$ and the fact that u_j and u_{j+1} are distinct words. A similar argument applies for the second of the possibilities unless for some $x \in c(L_h)\backslash\{x_1, x_2\}$ the letter x_1 appears in $\theta(x)$. In this case however, there is only one occurrence of x_1 to the right of x_2 in u_j and so x must be 1-occurring in L_h. The only remaining 1-occurring letter in L_h is x_4. However then for every $i \leq h$ there is an i' so that $\theta(y_i) \equiv x_2^{i'}$. Therefore $\theta(L_h) \equiv \theta(R_h)$, once again contradicting the fact that $u_j \not\equiv u_{j+1}$. So the only remaining possibility is that $\theta(x_1) \equiv x_1$ and $\theta(x_2) \equiv x_2^{pk+1}$. The same arguments show that $\theta(x_3) \equiv x_3$ and $\theta(x_4) \equiv x_4^{qk+1}$. In this case it is easily verified that $h = n$ and $\theta(y_i) \equiv y_i$ for all $i \leq n$. Thus $n \in M$ as required. □

We finish this section with a number of questions concerning semigroup varieties with uncountably many subvarieties.

QUESTION 4.2.14 *(i) Does* $\mathbf{A_2}$ *generate a variety with uncountably many subvarieties?*
(ii) Does $\mathbf{B_2}$ *generate a variety with uncountably many subvarieties (see also Question 4.2.12)?*

QUESTION 4.2.15 *Is there a finite WFB regular semigroup generating a variety with uncountably many subvarieties?*

Note that a negative answer to this question would imply a negative answer to both parts of Question 4.2.14 and enable a generalisation of Corollary 3.1.5.

QUESTION 4.2.16 *(i) What is the smallest finite semigroup (or monoid) generating a variety with uncountably many subvarieties?*
(ii) What is the smallest finitely based finite semigroup (or monoid) generating a variety with uncountably many subvarieties?

Two examples of seven element WFB semigroups were found above: $S(\{aba\})$ and that found in Example 4.2.9. In fact Theorem 4.1.2 and Theorem 4.2.5 enable the

construction of many such examples and a list of those found is presented in the appendix.

QUESTION 4.2.17 *Is there a finite monoid generating a variety with uncountably many subvarieties for which xyx is not an isoterm?*

A negative answer to this question would help improve the bounds for a solution to Question 4.2.16 as well as provide a partial solution to Question 4.1.17.

QUESTION 4.2.18 *Is the membership problem for the class of finite semigroups generating varieties with uncountably many subvarieties decidable?*

4.3 Connections with the uniform word problem

In several recent papers [11], [12], [13], [54], [96] (and in the doctoral thesis of B. Wells [95]) examples have been found of varieties, $\mathcal{V}$, with decidable word problem but undecidable uniform word problem (see the introduction to this chapter for a definition of these concepts). A second kind of example presented in the above papers are varieties $\mathcal{V}$ in which every finitely generated $\mathcal{V}$-free algebra has a decidable word problem but the equational theory of $\mathcal{V}$ is undecidable. Such a variety is said to be *psuedorecursive*. A further variation on these ideas are pseudorecursive varieties with decidable word problem (that is, pseudorecursive varieties in which *every* finitely presented algebra in the variety has a decidable word problem, not just the finitely generated free algebras); we will call such a variety *strongly pseudorecursive*. It is well known that the undecidability of the equational theory of a variety implies the undecidability of the uniform word problem for that variety. Thus a strongly pseudorecursive variety is also a variety of the first kind described above (trivially it is pseudorecursive as well). Examples of strongly pseudorecursive varieties are also presented in the above papers. Of particular interest is the example of Delić [13] which is a *finitely based* strongly pseudorecursive variety, although its basis is

quite complicated (see [12]). On the other hand the identity basis for the strongly pseudorecursive variety presented in [12] is infinite, but of quite a simple form.

To emphasise the subtleties of the above definitions it is worth considering a further property of pseudorecursive varieties. An identity involving at most n distinct letters is satisfied by a variety $\mathcal{V}$ if and only if it is satisfied by the $\mathcal{V}$-free algebra on n generators. If $\mathcal{V}$ is pseudorecursive then this fact can be algorithmically verified since the finitely generated $\mathcal{V}$-free algebras in $\mathcal{V}$ have a decidable word problem. Thus for any fixed $n \in \mathbb{N}$ there is an algorithm that determines if an identity in at most n distinct letters is satisfied by $\mathcal{V}$ but (since $\mathcal{V}$ is pseudorecursive) the equational theory of $\mathcal{V}$ is undecidable (this property is in fact an alternative definition of the concept of pseudorecursiveness)! Thus to show the decidability of an equational theory Σ it does not suffice to take an identity in n letters and construct an algorithm which determines if n is contained in Σ unless the actual algorithm does not depend on n. For further discussions of this nature, the reader is referred to [96].

As noted in [96] (see Remark 11.2.4) it is easy to establish the *existence* of (strongly) pseudorecursive varieties as follows. There are only countably many recursive sets of identities (sets of identities with decidable membership problem). Thus a variety with uncountably many subvarieties must contain uncountably many subvarieties with undecidable equational theory! In a locally finite variety all finitely presented algebras are finite (since they must be finitely generated if they are finitely presented) and a finite algebra (with finitely many operations) always has a decidable word problem (an algorithm is provided by the Cayley table for each of the operations of the algebra). Thus locally finite varieties with finitely many operations have decidable word problems and locally finite varieties with uncountably many subvarieties have uncountably many (strongly) pseudorecursive varieties. Of course this only establishes the existence of pseudorecursive varieties and does not give any explicitly.

In this section we show how to explicitly describe strongly pseudorecursive sub-

varieties of each of the locally finite varieties with uncountably many subvarieties which were found in the previous sections. We note (as in Remark 11.2.2 of [96]) that a finitely based locally finite variety has decidable equational theory and thus cannot be pseudorecursive. Consequently all the examples we will construct are NFB. They will however have a recursive basis of identities, that is they have a basis, Σ, of identities and there exists an algorithm that determines when an arbitrary identity is contained in Σ (the strongly pseudorecursive varieties described in [11], [12], [54], [95] and [96] are also NFB; a FB pseudorecursive variety with "no more than 350000 axioms" is found in [96] and a FB strongly pseudorecursive variety is found in [13]).

The method we will use is effectively the same as that used in many of the above papers: construct a locally finite variety with a recursive basis of identities but with undecidable equational theory. By the above comments this variety is strongly pseudorecursive. We initially formulate our results in a general, universal algebraic setting before applying them to the semigroup varieties of preceding sections. For further information regarding concepts of universal algebra see [9].

Recall that a primitive recursive function $\phi : \mathbb{N} \to \mathbb{N}$ is a function constructed in a basic way (namely by *composition* and *primitive recursion*) from certain fundamental functions on $\mathbb{N}$ (see [88] for a description of these fundamental functions and for a precise definition of a recursive function). Importantly, given ϕ and $n \in \mathbb{N}$ one can effectively compute $\phi(n)$. A subset $M \subseteq \mathbb{N}$ is said to be *recursively enumerable* if it is the empty set or it is the range of a recursive function and is said to be *recursive* if both M and $\mathbb{N} \backslash M$ are recursively enumerable. It is well known that there exist recursively enumerable sets that are not recursive (see [88] for example).

Let $\mathcal{V}$ be a variety of some type $\mathcal{F}$ and $Id(\mathcal{V})$ be the set of identities of $\mathcal{V}$ in some fixed countably infinite set of variables X. Let $F_{\mathcal{V}}(X)$ be the $\mathcal{V}$-free algebra generated by X. We now introduce the following definition.

DEFINITION 4.3.1 *An infinite list $W = \{w, w_1, w_2, \dots\}$ of type $\mathcal{F}$ terms from*

$T(X)$ *(the* term algebra *of type $\mathcal{F}$ over X) is said to be* strongly independent *with respect to $\mathcal{V}$ (or strongly independent with respect to Σ, a basis for $Id(\mathcal{V})$) if for every distinct pair of subsets P, $Q \subseteq \mathbb{N}$ the identities $\Sigma \cup \{w_n \approx w : n \in P\}$ and $\Sigma \cup \{w_n \approx w : n \in Q\}$ determine distinct subvarieties of $\mathcal{V}$.*

LEMMA 4.3.2 *Let $\phi : \mathbb{N} \to \mathbb{N}$ be a primitive recursive function such that $A = \{\phi(n) : n \in \mathbb{N}\}$ is a recursively enumerable but not recursive set and let $\mathcal{V}$ be a variety of type $\mathcal{F}$ algebras with recursive basis of identities Σ. If $W = \{w, w_1, w_2, \dots\}$ is a countably infinite recursive set of type $\mathcal{F}$ terms strongly independent with respect to $\mathcal{V}$ then the identities*

$$\Phi(\Sigma, W, \phi) = \Sigma \cup \{w_{2n} \approx w : n \in \mathbb{N}\} \cup \{w_{2n} \approx w_{2\phi(n)-1} : n \in \mathbb{N}\}$$

is a recursive basis for a subvariety $\mathcal{V}'$ of $\mathcal{V}$ with non recursive equational theory.

Proof: Firstly the identities $\Phi(\Sigma, W, \phi)$ form a basis for the identities of $\mathcal{V}'$ since W is strongly independent with respect to Σ. Secondly this basis is recursive: since Σ and $\{w_{2n} \approx w : n \in \mathbb{N}\}$ are recursive we need only check identities of the form $w_{2n} \approx w_{2m-1}$. Clearly such an identity is contained in $\Phi(\Sigma, W, \phi)$ if and only if m is the number $\phi(n)$, which can be effectively calculated.

The identities $\{w \approx w_{2\phi(n)-1}\}$ are easily seen to be a consequence of $\Phi(\Sigma, W, \phi)$. Since W is strongly independent with respect to $\mathcal{V}$, if M is any subset of $\mathbb{N}$ then $\Sigma \cup \{w_i \approx w : i \in M\} \vdash w_j \approx w$ if and only if $j \in M$. Thus $w_{2n-1} \approx w \in Id(\mathcal{V}')$ if and only if $n \in A$. Since the set A is not recursive, neither can be $Id(\mathcal{V}')$. □

For the remainder of this section we shall continue the notation $\Phi(\Sigma, W, \phi)$ from this lemma with the function ϕ a fixed primitive recursive function defining a recursively enumerable, nonrecursive set. We have the following.

COROLLARY 4.3.3 *If $\mathcal{V}$ is a locally finite variety with recursive basis Σ and W is a countably infinite, recursive set of terms W that is strongly independent with respect to $\mathcal{V}$ then the variety $\mathcal{V}'$ defined by the identities $\Phi(\Sigma, W, \phi)$ is a recursively based, strongly pseudorecursive variety.*

In the previous sections of this chapter we found varieties $\mathcal{V}$ with infinite, recursive lists of words W that are strongly independent in $\mathcal{V}$. By Corollary 4.3.3 if Σ is a recursive basis for $\mathcal{V}$ then $\Phi(\Sigma, W, \phi)$ is a recursively based, strongly pseudorecursive variety of semigroups. Since (as noted above) a finitely based locally finite variety has decidable equational theory, HFB varieties have no pseudorecursive subvarieties. Thus, by Corollary 4.1.3 for example, we have the following result.

COROLLARY 4.3.4 *If $\mathcal{V}$ is a locally finite variety with recursive basis Σ and $\mathcal{V}$ is generated by a monoid of index more than two, $\mathcal{V}$ has a pseudorecursive subvariety if and only if $\mathcal{V}$ is not hereditarily finitely based and if and only if xyx is an isoterm for $\mathcal{V}$. If xyx is an isoterm for $\mathcal{V}$ then the variety described by $\Phi(\Sigma, \mathcal{W}, \phi)$ where $\mathcal{W}$ is the list $\{(y_1x_1x_2x_3x_4y_1)y_2x_5y_2y_3x_6y_3 \dots y_{n-1}x_{n+2}y_{n-1}(y_nx_{n+3}x_{n+4}x_{n+5}x_{n+6}y_n) : n \in \mathbb{N}\}$ is a strongly pseudorecursive subvariety of $\mathcal{V}$.*

There are corresponding corollaries of this kind for all results from preceding sections concerning locally finite varieties with uncountably many subvarieties.

A simple example is the following.

EXAMPLE 4.3.5 *The variety defined by $\Phi(\Sigma, \mathcal{W}, \phi)$ where Σ is the set*

$$\{(xy)z \approx x(yz),\ x^3 \approx x^4,$$

$$xy_1xy_2x \approx y_1y_2xxx,\ xxy_1x \approx y_1xxx,\ xy_1xx \approx y_1xxx,\ xxxy_1 \approx y_1xxx\}$$

and $\mathcal{W}$ is the list of words in Corollary 4.3.4 is strongly pseudorecursive.

Proof: The set of identities Σ is obviously equivalent to the identities $\mathcal{A}_3$ which by Corollary 2.2.3 form a basis of the identities of the semigroup $S(\{abab, aabb, abba\})$. Now Theorem 4.1.8 and the proof of Theorem 4.1.2 implies that the words $\mathcal{W}$ of the example are strongly independent with respect to $\mathcal{V}(S(\{abab, aabb, abba\}))$ and the result then follows by Corollary 4.3.3.

The basis of the variety in Example 4.3.5 is obtained by adjoining the six identities to an infinite set of identities (though this infinite set contains identities of

two slightly different forms). We note that the simple example of a strongly pseudorecursive groupoid variety in [12] also has six identities adjoined to an infinite system.

The existence of a (strongly) pseudorecursive variety of groups is noted in [96] and the possible existence of a recursively based example is an open question in [96] (see Remark 11.4.2). We now show how some existing results can be combined with Corollary 4.3.3 above to provide such an example. Let $[x, y]$ be the group theoretic commutator $x^{-1}y^{-1}xy$ and for $n \geq 3$, let $[x_1, x_2, \ldots, x_n] = [[x_1, \ldots x_{n-1}], x_n]$. In [94] it is shown that the group words T given by

$$\{x^{16},\ [[y_1, y_2, y_3], [x_1, x_2], [x_3, x_4], \ldots, [x_{2i-1}, x_{2i}], [y_1, y_2, y_3]] : i \in IN\}$$

are strongly independent with respect to the (locally finite) group variety $\mathcal{V}$ defined within the variety of groups by the identity $w \approx 1$ where w is the word $x^{16}[[z_1, z_2, z_3], [z_4, z_5, z_6], [z_7, z_8]]$ (a similar result is found in [61]). Therefore Corollary 4.3.3 implies that $\Phi = \Phi(\{(xy)z \approx x(yz),\ x1 \approx x,\ xx^{-1} \approx x^{-1}x \approx w \approx 1\}, T, \phi)$ determines a strongly pseudorecursive variety of groups.

Chapter 5

Some undecidable embedding problems for finite semigroups.

In this chapter we consider a number of embedding problems which have no algorithmic solution. In each case we use a method that first appeared in [25]. Roughly speaking (a precise description will be given in the following section) we consider an arbitrary partially defined finite group **G** and from it construct a new structure **S** with the property that **S** is embeddable in a semigroup of the desired form exactly when the original partial group can be embedded in a group. It follows from a result of T. Evans (see Connection 2.2 of [39]) that the set of finite "partial groups" embeddable in a finite group, or a group, is not recursive and therefore the set of structures embeddable in finite semigroups or semigroups with the desired property also is not recursive. This method appears to be extremely useful in showing various embedding problems to be undecidable and has been used in a number of recent papers: [25], [30], [45], [46] and [76] (the second paper in this list concerns the results in Sections 5.2 and 5.3.2 to follow).

5.1 Preliminaries

A number of preliminary results and definitions are required before we prove the main results of this chapter. The following concept was introduced by M. Sapir and is a useful tool in "transcribing" the structure of a partial group into an appropriate semigroup structure.

DEFINITION 5.1.1 *(M. Sapir) A split system is a triple of sets (A, B, C) with an associated operation $A \times B \to C$. An embedding of a split system into a semigroup $\mathbf{S}$ is a triple of maps (i, j, k) such that the maps $i : A \to \mathbf{S}$, $j : B \to \mathbf{S}$ and $k : C \to \mathbf{S}$ are injective and $i(a)j(b) = k(ab)$, for each $a \in A$ and $b \in B$.*

On occasions the generality of this concept is unnecessary and it is convenient to instead use a simplified notion as follows.

DEFINITION 5.1.2 *A split pair is a pair of sets (A, B) with an associated operation $A \times A \to B$. An embedding of a split pair into a semigroup $\mathbf{S}$ is a pair of maps (j, k) such that the maps $j : A \to \mathbf{S}$ and $k : B \to \mathbf{S}$ are injective and $j(a)j(b) = k(ab)$, for each $a, b \in A$.*

By a partial group $\mathbf{G}$ we will mean a set with an element 1 and a partially defined binary operation such that for every $x \in \mathbf{G}$, $1x = x1 = x$ and if both $(xy)z$ and $x(yz)$ are defined then they are equal. The following definition appears in [25]. (For the purposes of this definition it is convenient to make a distinction between a semigroup (or partial semigroup) $\mathbf{S}$ and its universe S.)

DEFINITION 5.1.3 *Let $\mathbf{G_0}$ and $\mathbf{G}$ be partial groups such that $\mathbf{G_0}$ is embedded in $\mathbf{G}$. For each $i = 0, 1, 2, \ldots$, let G_0^i be the subset of the universe of $\mathbf{G}$ defined as follows: $G_0^0 = \{1\}$ (the identity element), $G_0^1 = G_0$, $G_0^{i+1} = G_0^i G_0$. Then for $k \geq 2$, the partial group $\mathbf{G}$ is an extension of rank k of $\mathbf{G_0}$ if and only if*
(i) $G = \bigcup_{i=0}^{k} G_0^i$,
(ii) for every pair of positive integers i, j with $i + j \leq k$ and every pair of elements

$x \in G_0^i$, $y \in G_0^j$, the product xy exists and is contained in G_0^{i+j},
(iii) if $i+j > k$ and $x \in G_0^i \setminus G_0^{i-1}$, $y \in G_0^j \setminus G_0^{j-1}$ then the product xy is not defined,
(iv) if $i+j+l \leq k$ and $x \in G_0^i$, $y \in G_0^j$, $z \in G_0^l$, then $(xy)z$ and $x(yz)$ are defined and equal,
(v) for $f, g, h \in G$, if $fg = fh$ or $gf = hf$, then $g = h$.

From Connection 2.2 in [39] we have that the unsolvability of the uniform word problems in the pseudovariety of all groups (which, of course, is also a variety) and in the pseudovariety of finite groups imply that the problem of determining whether a finite partial group is embeddable in a group or in a finite group is undecidable. A group $\mathbf{H}$ can be viewed trivially as an extension of arbitrary rank of itself. So for every k, a partial group $\mathbf{G}$ is embeddable in a group (or a finite group), $\mathbf{H}$, if and only if there is an extension of rank k of $\mathbf{G}$ that is embeddable in $\mathbf{H}$. If the problem of determining whether or not an extension of rank k of a partial group is embeddable in a group (or a finite group) is decidable then we would obtain the following algorithm for determining when an arbitrary finite partial group $\mathbf{G}$ is embeddable in a group (or a finite group), contradicting the fact that this second problem is undecidable:

1. Construct all extensions of rank k of $\mathbf{G}$ (there are only finitely many and they can be effectively listed);

2. If one of the extensions of rank k is embeddable in a group (or a finite group), $\mathbf{H}$, then $\mathbf{G}$ is embeddable in $\mathbf{H}$. Otherwise $\mathbf{G}$ is not embeddable in a group (or a finite group).

We therefore have the following lemma:

LEMMA 5.1.4 *[25] The problem of determining whether or not an extension of rank k of a partial group is embeddable in a group or in a finite group is undecidable.*

DEFINITION 5.1.5 *An $\mathcal{H}$-embedding of a split system (A, B, C) (or split pair (A, B)) is an embedding (i, j, k) (or (i, j) respectively) of (A, B, C) ((A, B) respec-*

tively) into a semigroup $\mathbf{S}$ *so that* $i(A)$, $j(B)$ *and* $k(C)$ *(or* $i(A)$, $j(B)$ *respectively) lie within* $\mathcal{H}$ *classes of* $\mathbf{S}$.

For a given $\mathbf{G}$, an extension of rank 2 of a finite partial group $\mathbf{G_0}$, we can construct an *associated split system* (A, B, C) where $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ are disjoint copies of $\mathbf{G_0}$, $C = \{c_1, \ldots, c_m\}$ a copy of $\mathbf{G}$, and with operation $a_i b_j = c_k$ whenever $g_i g_j = g_k$ in $\mathbf{G}$. In an analogous way we can construct an associated split pair (A, C) by replacing the requirement that the sets A and B are disjoint with the requirement that they are identical. It is clear that an embedding of a split pair constructed in this way determines an embedding of the original split system. Furthermore any embedding θ of $\mathbf{G}$ into a group determines a natural embedding (i, j) of the split pair (A, C) (with $i(a_k) = j(c_k) = \theta(g_k)$). Part (i) of the next lemma is Lemma 7 of [76] and part (ii) follows from the arguments above.

LEMMA 5.1.6 *(i) (M. Sapir) Let* (A, B, C) *be the split system associated with* $\mathbf{G}$, *an extension of rank 2 of a finite partial group* $\mathbf{G_0}$. *There is an* $\mathcal{H}$*-embedding* (i, j, k) *of* (A, B, C) *into a semigroup* $\mathbf{S}$ *if and only if* $\mathbf{G}$ *is embeddable in a subgroup of* $\mathbf{S}$. *(ii) Let* (A, B) *be the split system associated with* $\mathbf{G}$, *an extension of rank 2 of a finite partial group* $\mathbf{G_0}$. *There is an* $\mathcal{H}$*-embedding* (i, j) *of* (A, B) *into a semigroup* $\mathbf{S}$ *if and only if* $\mathbf{G}$ *is embeddable in a subgroup of* $\mathbf{S}$.

In the following sections we will be constructing semigroups (or related structures) whose structure is determined by certain extensions of rank k of partial groups. Because of Lemma 5.1.6, the use of split systems (and in the following section, split pairs) is helpful in simplifying arguments concerned with connecting the embeddability properties of partial groups with the desired embedding properties of our constructions.

5.2 Potentially $\mathcal{H}$-embeddable subsets

In this section we examine a natural embedding problem concerned with the $\mathcal{H}$ relation of Green. Let $\mathcal{U}$ represent one of Green's relations on a semigroup $\mathbf{S}$. From the definitions of Green's relations (see Section 1.3.2) it is easy to determine when given a subset A of a semigroup $\mathbf{S}$ whether or not A lies in $\mathcal{U}$-class of $\mathbf{S}$ Furthermore, if A is a $\mathcal{U}^{\mathbf{S}}$-related subset of $\mathbf{S}$ and θ is an embedding of $\mathbf{S}$ into a semigroup $\mathbf{T}$ then $\theta(A)$ is a $\mathcal{U}^{\mathbf{T}}$-related subset of $\mathbf{T}$. On the other hand the restriction of a $\mathcal{U}^{\mathbf{T}}$-class of a semigroup $\mathbf{T}$ to some subsemigroup $\mathbf{S}$ need not be a $\mathcal{U}^{\mathbf{S}}$-class.

DEFINITION 5.2.1 *If $\mathbf{S}$ is a finite semigroup and $A \subseteq \mathbf{S} \times \mathbf{S}$ then we say A is potentially[1] $\mathcal{U}$-related if $A \subseteq \mathcal{U}^{\mathbf{T}}$ for some supersemigroup $\mathbf{T}$ containing $\mathbf{S}$. If $\mathbf{T}$ can be chosen from a particular class $\mathcal{K}$ of semigroups (the class of finite semigroups for example) then we say A is potentially $\mathcal{U}$-related in $\mathcal{K}$. If $A \subseteq \mathbf{S}$ then we say that A is potentially $\mathcal{U}$-embeddable in a class $\mathcal{K}$ if $A \times A$ is potentially $\mathcal{U}$-related in $\mathcal{K}$.*

Note that if there exists an algorithm determining for an arbitrary semigroup $\mathbf{S}$ whether or not a given finite subset of $\mathbf{S} \times \mathbf{S}$ is potentially $\mathcal{U}$-related then there certainly exists an algorithm determining if a given finite subset of $\mathbf{S}$ is potentially $\mathcal{U}$-embeddable.

Define the following relations on a semigroup $\mathbf{S}$:

$$\mathcal{L}^* = \{(a,b) : ax = ay \Leftrightarrow bx = by \ \ \forall x, y \in \mathbf{S}^1\},$$

$$\mathcal{R}^* = \{(a,b) : xa = ya \Leftrightarrow xb = yb \ \ \forall x, y \in \mathbf{S}^1\},$$

$$\mathcal{H}^* = \mathcal{L}^* \wedge \mathcal{R}^*.$$

We have the following well known result (for example, see [21], [48] or [62]).

[1] In some of the literature ([30] and [76] for example), the word "eventually" is used here instead of "potentially".

LEMMA 5.2.2 *If* $\mathbf{S}$ *is a semigroup then a subset* $A \subseteq \mathbf{S} \times \mathbf{S}$ *is potentially* $\mathcal{L}$*-related (respectively potentially* $\mathcal{R}$*-related) if and only if* $A \subseteq \mathcal{L}^*$ *(respectively* $A \subseteq \mathcal{R}^*$*). Furthermore if* $\mathbf{S}$ *is finite, then a subset* $A \subseteq \mathbf{S} \times \mathbf{S}$ *is potentially* $\mathcal{L}$*-related (respectively potentially* $\mathcal{R}$*-related) if and only if it is potentially* $\mathcal{L}$*-related within the class of finite semigroups (respectively potentially* $\mathcal{R}$*-related within the class of finite semigroups).*

This lemma works for $\mathcal{L}^*$ (resp. $\mathcal{R}^*$) because of the left (respectively right) regular representation of $\mathbf{S}$ by inner left (respectively right) translations on the set S^1. There is no natural analogue of this for the $\mathcal{H}$-relation.

Lemma 5.2.2 provides a simple algorithm for testing whether a given subset of a finite semigroup is potentially $\mathcal{L}$-embeddable (or potentially $\mathcal{R}$-embeddable). In [76] however, M. V. Sapir has shown that the problem of determining, for two disjoint subsets A, B of a finite semigroup $\mathbf{S}$, whether or not $(A \times A) \cup (B \times B)$ is potentially $\mathcal{H}$-related is undecidable. This, along with Lemma 5.2.2, implies the existence of a finite semigroup $\mathbf{S}$ and a subset $(A \times A) \cup (B \times B)$ of $\mathbf{S} \times \mathbf{S}$ for which $(A \times A) \cup (B \times B) \subseteq \mathcal{H}^*$ but are not potentially $\mathcal{H}$-related (Corollary 1 of [76]). The main aims of this section will be to provide examples of such semigroups and to prove the following extension of the results in [76].

THEOREM 5.2.3 *The problem of determining whether or not a subset* A *of a finite semigroup* $\mathbf{S}$ *is potentially* $\mathcal{H}$*-embeddable in the class of finite semigroups or in the class of all semigroups is undecidable.*

Problem 1 of [76] asks if there is an algorithm for determining whether a subset A of a finite semigroup $\mathbf{S}$ is potentially $\mathcal{H}$-embedded in the class of finite semigroups. Theorem 5.2.3 answers this in the negative. It is also remarked in [76] that there is an algorithm for determining whether or not a subset A of a finite semigroup $\mathbf{S}$ is potentially $\mathcal{H}$-embedded in the class of all semigroups. This statement is not proved in [76] and in fact Theorem 5.2.3 of this thesis shows that it is not true.

For the arguments to follow, let $\mathbf{G}$ always denote an extension of rank 2 of a partial group $\mathbf{G_0}$ with the elements of $\mathbf{G_0}$ labeled $\{g_1, g_2, \ldots, g_n\}$ and such that g_1 is the identity element. Let the remaining elements of $\mathbf{G}$ be labeled $\{g_{n+1}, \ldots, g_m\}$.

DEFINITION 5.2.4 *For the split pair (A, B) associated with $\mathbf{G}$, an extension of rank 2 of a partial group $\mathbf{G_0}$, define $\mathbf{S_{(G,G_0)}}$ to be the semigroup whose universe is the set $\{0\} \cup A \cup B$ and with multiplication $a_i \cdot a_j = b_k$ if $a_i a_j = b_k$ in (A, B) and 0 otherwise.*

The groupoid $\mathbf{S_{(G,G_0)}}$ is a semigroup, since the product of any three elements in $\mathbf{S_{(G,G_0)}}$ is zero (that is, $\mathbf{S_{(G,G_0)}}$ is 3-nilpotent).

DEFINITION 5.2.5 *If $\mathbf{C}$ is a group then define $\overline{\mathbf{C}}$ as the semigroup whose universe is $\mathbf{C} \cup A_c \cup B_c \cup \{0\}$, where A_c and B_c are disjoint copies of the universe of $\mathbf{C}$ and with multiplication (for $a_i \in A_c$, $b_i \in B_c$, $c_i \in \mathbf{C}$, and where x_i is one of a_i, b_i, or c_i)*

$$a_i \cdot a_j = b_k, \text{ if } c_i c_j = c_k \text{ in } \mathbf{C}$$

$$x_i \cdot c_j = c_i \cdot x_j = x_k, \text{ if } c_i c_j = c_k \text{ in } \mathbf{C}$$

and all other products take the value 0.

Now $\overline{\mathbf{C}}$ is a semigroup since the subscripts of the elements behave as in the group $\mathbf{C}$ and the letter names of the elements behave according to the following table:

$\cdot$	0	A_c	B_c	$\mathbf{C}$
0	0	0	0	0
A_c	0	B_c	0	A_c
B_c	0	0	0	B_c
$\mathbf{C}$	0	A_c	B_c	$\mathbf{C}$

which is a commutative, 3-nilpotent semigroup with adjoined identity element, $\mathbf{C}$ (indeed, $\overline{\mathbf{C}}$ is an extension of this semigroup). Note that since $\mathbf{C}$ is a group, the $\mathcal{H}^{\overline{\mathbf{C}}}$-classes of $\overline{\mathbf{C}}$ are $\{0\}$, A_c, B_c, and $\mathbf{C}$.

Theorem 5.2.3 follows from the following lemma and Lemma 5.1.4 part (ii).

LEMMA 5.2.6 *Let (A, B) be the split pair associated with $\mathbf{G}$, an extension of rank 2 of a partial group $\mathbf{G_0}$. The subset A of $\mathbf{S_{(G,G_0)}}$ is potentially $\mathcal{H}$-embeddable in the class of semigroups (or finite semigroups) if and only if $\mathbf{G}$ is embeddable in a group (or a finite group).*

Proof: Suppose θ is an embedding of $\mathbf{G}$ into a group $\mathbf{C}$, with the elements of $\mathbf{C}$ labeled so that $\theta(g_i) = c_i$. Then $\theta' : \mathbf{S_{(G,G_0)}} \to \overline{\mathbf{C}}$ defined by

$$\theta'(a_i) = a_i \in A_c,\ \theta'(b_i) = b_i \in B_c,\ \theta'(0) = 0$$

is an embedding of $\mathbf{S_{(G,G_0)}}$ in $\overline{\mathbf{C}}$ which sends A to a subset of the $\mathcal{H}^{\overline{\mathbf{C}}}$-class A_c.

So now assume that $\mathbf{S_{(G,G_0)}}$ is the subsemigroup of a bigger semigroup $\mathbf{T}$, in which A lies in an $\mathcal{H}^{\mathbf{T}}$-class, H_A. We may assume that $\mathbf{T}$ is regular, since every (finite) semigroup can be embedded into a (finite) regular semigroup, and its $\mathcal{H}$-classes will still be within $\mathcal{H}$-classes of the regular semigroup. Now for every $g_i, g_j \in \mathbf{G_0}$, whenever $xa_i = a_j$ and $ya_j = a_i$, for some $x, y \in \mathbf{T}$ we have $xa_ia_1 = a_ja_1$ and $ya_ja_1 = a_ia_1$, or $xb_i = b_j$ and $yb_j = b_i$, so therefore $b_i\mathcal{L}^{\mathbf{T}}b_j$. Similarly, $b_i\mathcal{R}^{\mathbf{T}}b_j$, and thus, $b_i\mathcal{H}^{\mathbf{T}}b_j$. For $b_k \in B$, with $g_k \notin \mathbf{G_0}$, we can find (by the definition of A) $a_i, a_j \in A$ with $a_ia_j = b_k$. Since $A \subseteq H_A$, there exists $x, y \in \mathbf{T^1}$ with $xa_i = a_1,\ ya_1 = a_i$. So

$$xb_k = xa_ia_j = a_1a_j = b_j \text{ and } yb_j = ya_1a_j = a_ia_j = b_k$$

and hence it follows that $b_k\mathcal{L}^{\mathbf{T}}b_j$. Similarly $b_k\mathcal{R}^{\mathbf{T}}b_i$ and since $b_i\mathcal{H}^{\mathbf{T}}b_j$, we have shown that B is contained in an $\mathcal{H}^{\mathbf{T}}$-class, H_B.

We can now use Lemma 5.1.6 (ii). By construction, there is a natural embedding (i, j) of the split pair (A, B) into $\mathbf{S_{(G,G_0)}}$. Since $\mathbf{S_{(G,G_0)}}$ is embedded in a semigroup

$\mathbf{T}$ where A and B lie in $\mathcal{H}^{\mathbf{T}}$ classes this embedding can be extended to an $\mathcal{H}$ embedding of (A, B) and therefore by Lemma 5.1.6 the extension $\mathbf{G}$ of rank 2 of the partial group $\mathbf{G_0}$ is embeddable in a subgroup of $\mathbf{T}$ as required. □

Theorem 5.2.3 is proved. □

We now turn our attention to the construction of $\mathcal{H}^*$-related subsets of semigroups that are not potentially $\mathcal{H}$-embeddable. In [76] it is proved that there exists a finite semigroup $\mathbf{S}$ with a subset A of $\mathbf{S} \times \mathbf{S}$ that satisfies $A \subseteq \mathcal{H}^*$ but which is not potentially $\mathcal{H}$-related. Theorem 5.2.3 implies the existence of a finite semigroup for which there is an $\mathcal{H}^*$-class that is not potentially $\mathcal{H}$-embeddable. Such an example is not presented in [76] nor seems to have been published elsewhere. By Lemma 5.2.6 the subset A of the semigroup $\mathbf{S}_{(\mathbf{G},\mathbf{G_0})}$ is potentially $\mathcal{H}$-embeddable if and only if $\mathbf{G}$ is embeddable in a group. Thus to find the desired example it suffices to find an extension of rank 2 of a partial group that is not embeddable in any group.

EXAMPLE 5.2.7 *Consider the eight element semigroup* $\mathbf{S}_{(\mathbf{G},\mathbf{G_0})}$ *where* $\mathbf{G_0}$ *is the partial group defined by:*

$\cdot$	g_1	g_2	g_3
g_1	g_1	g_2	g_3
g_2	g_2	g_3	
g_3	g_3		g_2

$\mathbf{G}$ *is an extension of rank 2 of* $\mathbf{G_0}$ *defined by:*

$\cdot$	g_1	g_2	g_3	g_4
g_1	g_1	g_2	g_3	g_4
g_2	g_2	g_3	g_4	
g_3	g_3	g_4	g_2	
g_4	g_4			

Then the three element subset A of $\mathbf{S}_{(\mathbf{G_0},\mathbf{G})}$ is an $\mathcal{H}^*$ class of $\mathbf{S}_{(\mathbf{G_0},\mathbf{G})}$ that is not potentially $\mathcal{H}$-embeddable in the class of semigroups (and therefore in the class of finite semigroups).

Proof: In $\mathbf{G}$ we have $(g_2g_2)(g_2g_2) = g_3g_3 = g_2 = g_2g_1$ and $g_2(g_2(g_2g_2)) = g_2g_4$ so therefore $g_2g_1 = g_2g_4$, a property not satisfied by any group. Thus $\mathbf{G}$ is not embeddable in a group and so by Lemma 5.2.6, the subset A is not potentially $\mathcal{H}$-embeddable. It is easily verified that A is an $\mathcal{H}^*$-class of $\mathbf{S}_{(\mathbf{G},\mathbf{G_0})}$. □

While Lemma 5.2.6 shows that any extension of rank 2 of a partial group not embeddable in a group will give rise to a semigroup with a subset that is not $\mathcal{H}$-embeddable, it is a very simple and routine exercise to show that any 3 element extension of rank 2 of a partial group is always embeddable in a group and so no smaller examples can be obtained by exactly the methods used above. This fact also makes it impossible to use the above method to construct semigroups with an $\mathcal{H}^*$-related pair that are not potentially $\mathcal{H}$-related. The following 3-nilpotent semigroup $\mathbf{S}$ shows that such examples nevertheless exist.

EXAMPLE 5.2.8 *Let $\mathbf{S}$ be the semigroup given by the following cayley table*

$\cdot$	0	a_1	a_2	b_1	b_2	c_1	c_2	c_3
0	0	0	0	0	0	0	0	0
a_1	0	0	0	c_1	c_3	0	0	0
a_2	0	0	0	c_2	c_3	0	0	0
b_1	0	c_1	c_3	0	0	0	0	0
b_2	0	c_2	c_2	0	0	0	0	0
c_1	0	0	0	0	0	0	0	0
c_2	0	0	0	0	0	0	0	0
c_3	0	0	0	0	0	0	0	0

Then the set $A = \{a_1, a_2\}$ *in* $\mathbf{S}$ *is an* $\mathcal{H}^*$*-class of* $\mathbf{S}$ *but is not potentially* $\mathcal{H}$*-embeddable.*

Proof: A is an $\mathcal{L}^*$-class of $\mathbf{S}$, since for $i \in \{1,2\}$, $a_i x = a_i y$, for $x, y \in \mathbf{S}^1$, $x \neq y$ if and only if both x and y are contained in $\{0, a_1, a_2, c_1, c_2, c_3\}$. Likewise, A is an $\mathcal{R}^*$-class and therefore an $\mathcal{H}^*$-class.

Now let $\mathbf{T}$ be any semigroup in which $\mathbf{S}$ can be embedded so that A is $\mathcal{L}^{\mathbf{T}}$-related. So there is an $x \in \mathbf{T}$ such that $xa_1 = a_2$ (of course we may assume that x is not the identity element of $\mathbf{T}$ since $a_1 \neq a_2$). Therefore,

$$(xb_1)a_1 = x(b_1a_1) = xc_1 = xa_1b_1 = a_2b_1 = c_2 = b_2a_1.$$

However

$$(xb_1)a_2 = x(b_1a_2) = xc_3 = xa_1b_2 = a_2b_2 = c_3 \neq c_2 = b_2a_2.$$

So therefore A, as a subset of $\mathbf{T}$ is not $\mathcal{R}^*$-related. That is, whenever A is $\mathcal{L}$-related in some embedding semigroup, it is neither $\mathcal{R}$-related nor potentially $\mathcal{R}$-related in that semigroup. □

Infinite examples consisting of single $\mathcal{H}^*$-classes that are not potentially $\mathcal{H}$-related are also known. For example J. Fountain has noted (see comment in [76]) that any cancellative semigroup not embeddable in a group is $\mathcal{H}^*$-related but not potentially $\mathcal{H}$-embeddable (see [10] for such an example by A. Malcev). By taking the 0-direct join of any of the above examples with an infinite null semigroup one obtains other examples of infinite semigroups with $\mathcal{H}^*$-related, not eventually $\mathcal{H}$-embeddable subsets. On the other hand, it is a simple task to prove that a finite semigroup for which $\mathcal{H}^*$ is the universal relation (as in the infinite examples suggested by Fountain) is a group.

5.3 Potentially $\mathcal{L} - \mathcal{R}$-embeddable subsets

Using essentially the same method as in the previous section we now prove a variation on Theorem 5.2.3. For a subset A of a semigroup $\mathbf{S}$ to be potentially $\mathcal{H}$-embeddable there must be an embedding semigroup $\mathbf{T}$ so that A is simultaneously $\mathcal{L}^{\mathbf{T}}$ and $\mathcal{R}^{\mathbf{T}}$-related. We now replace the notion of a subset being potentially $\mathcal{H}$-embeddable with a similar but possibly weaker condition on pairs of disjoint subsets.

DEFINITION 5.3.1 *If A and B are disjoint subsets of a finite semigroup $\mathbf{S}$ then the pair $[A, B]$ is potentially $\mathcal{R} - \mathcal{L}$-embeddable if there is a supersemigroup $\mathbf{T}$ containing $\mathbf{S}$ in which A is contained in a $\mathcal{R}^{\mathbf{T}}$-class and B is contained in an $\mathcal{L}^{\mathbf{T}}$-class. $[A, B]$ is potentially $\mathcal{R} - \mathcal{L}$-embeddable in $\mathcal{K}$ if $\mathbf{T}$ can be chosen from a particular class $\mathcal{K}$ of semigroups.*

We now prove an analogous result to Theorem 5.2.3 concerning potentially $\mathcal{R}-\mathcal{L}$-embeddable pairs of disjoint subsets.

THEOREM 5.3.2 *The problem of determining for two disjoint subsets A and B of a finite semigroup $\mathbf{S}$ whether $[A, B]$ is potentially $\mathcal{R} - \mathcal{L}$-embeddable in the class of all semigroups and in the class of finite semigroups is undecidable.*

As before, for all arguments to follow in this section, we will assume that $\mathbf{G}$ is an extension of rank 2 of a partial group $\mathbf{G_0}$.

For the purposes of the following definition it is again convenient to make a distinction between a partial semigroup and its universe.

DEFINITION 5.3.3 *Let $\mathbf{G_2}$ be an extension of rank 3 of $\mathbf{G}$, and let G_1 be the set theoretic union $G^2 \cup G$. (Here and for the rest of this definition we assume the multiplication of $\mathbf{G_2}$ on the subsets G_i of G_2. This means, for example, that the set $G^2 \backslash G$ may be non empty.) Let A, B, C, D be disjoint copies of the sets G_0, G, G_1, G_2 respectively. Then define $\mathbf{S}_{(\mathbf{G},\mathbf{G_0},\mathbf{G_1},\mathbf{G_2})}$ to be the semigroup whose universe is $A \cup$*

$B \cup C \cup D \cup \{0\}$ *and has the following operation:*

$$a_i a_j = b_k, \text{ whenever } g_i, g_j \in G_0, \text{ and } g_i g_j = g_k \in G,$$

$$a_i b_j = b_i a_j = c_k, \text{ whenever } g_i g_j = g_k \in G_1 \text{ and } g_i \in G_0,\ g_j \in G \text{ or reverse},$$

$$a_i c_j = c_i a_j = d_k, \text{ whenever } g_i g_j = g_k \in G_1 \text{ and } g_i \in G_0,\ g_j \in G_1 \text{ or reverse},$$

$$b_i b_j = d_k, \text{ whenever } g_i, g_j \in G \text{ and } g_i g_j = g_k \in G_2,$$

$$0, \text{ otherwise.}$$

Note that $\mathbf{S}_{(\mathbf{G},\mathbf{G_0},\mathbf{G_1},\mathbf{G_2})}$ is a semigroup since the subscripts of elements behave according to the extension of rank 3 of $\mathbf{G}$ which is associative, and the letter names behave according to the 5-nilpotent semigroup

.	0	A	B	C	D
0	0	0	0	0	0
A	0	B	C	D	0
B	0	C	D	0	0
C	0	D	0	0	0
D	0	0	0	0	0

for which associativity can be routinely verified.

Theorem 5.3.2 now follows from Lemma 5.1.4 and the following lemma.

LEMMA 5.3.4 *Let* $\mathbf{G}$ *be an extension of rank 2 of a partial group* $\mathbf{G_0}$. *Then* $\mathbf{G}$ *is embeddable in a group (or a finite group) if and only if there exists an extension* $\mathbf{G_2}$ *of rank 3 of* $\mathbf{G}$ *such that for the subsets* A *and* B *of the semigroup* $\mathbf{S}_{(\mathbf{G},\mathbf{G_0},\mathbf{G_1},\mathbf{G_2})}$ *(with* G_1 *appropriately defined),* $[A, B]$ *is potentially* $\mathcal{R} - \mathcal{L}$*-embeddable in the class of semigroups (or in the class of finite semigroups respectively).*

Proof: Firstly assume $\mathbf{G}$ is embeddable in a group $\mathbf{H}$ and $\mathbf{G_2}$ is an extension of rank 3 of $\mathbf{G}$ that is compatible with the multiplication of $\mathbf{H}$ (that is, $\mathbf{G_2}$ is embeddable in $\mathbf{H}$). Then by adjoining an identity element, 1, to the table above and then constructing a new semigroup $\mathbf{T}$ by replacing the letters $A, B, C, D, 1$ with disjoint copies of the group $\mathbf{H}$ as in Definition 5.2.5, it is quickly seen that $\mathbf{S_{(G,G_0,G_1,G_2)}}$ is embedded in $\mathbf{T}$ such that all of the sets $A, B, C, D, \{0\}$ lie in $\mathcal{H}^{\mathbf{T}}$-classes. So certainly $[A, B]$ is potentially $\mathcal{R} - \mathcal{L}$-embeddable. Notice also that $\mathbf{T}$ is finite if and only if $\mathbf{H}$ is finite.

So now assume there is an extension $\mathbf{G_2}$ of rank 3 of $\mathbf{G}$ such that the semigroup $\mathbf{S_{(G,G_0,G_1,G_2)}}$ (with G_1 defined as before) is embedded in a semigroup $\mathbf{T}$ in which $[A, B]$ is $\mathcal{R} - \mathcal{L}$-embedded. Proceeding as in the proof of Lemma 5.2.6 from the last section, we have that A being $\mathcal{R}^{\mathbf{T}}$-related implies that B is $\mathcal{R}^{\mathbf{T}}$ -related. But B is $\mathcal{L}^{\mathbf{T}}$-related by our assumption, so therefore B is potentially $\mathcal{H}$-embeddable. We now show that there is an extension $\mathbf{G_3}$ of rank 2 of $\mathbf{G}$ (itself an extension of rank 2 of $\mathbf{G_0}$) for which the semigroup $\mathbf{S_{(G_3,G)}}$ is the subsemigroup of $\mathbf{S_{(G_0,G,G_1,G_2)}}$ generated by the set B and therefore by Lemma 5.2.6, $\mathbf{G_3}$, hence $\mathbf{G}$, is embeddable in a group (and if $\mathbf{T}$ is finite, then $\mathbf{G}$ is embeddable in a finite group).

Let $D' = \{d_k \in D : b_i b_j = d_k\}$. Consider the extension $\mathbf{G_3}$ of rank 2 of $\mathbf{G}$ whose universe is the set G_1, and whose multiplication is $g_i g_j = g_k$, if $g_i, g_j \in \mathbf{G}$ and $g_i g_j = g_k$ in the extension of rank 3 $\mathbf{G_2}$; $g_i g_1 = g_1 g_i = g_i$, if $g_i \in G_1$; and undefined otherwise. This is a "sub partial group" of $\mathbf{G_2}$ and therefore the semigroup $\mathbf{S_{(G_3,G)}}$ is isomorphic to the subsemigroup of $\mathbf{S_{(G,G_0,G_1,G_2)}}$ on the set $\{0\} \cup B \cup D'$. Since B is $\mathcal{H}$ related in $\mathbf{T}$, Lemma 5.2.6 applies and so $\mathbf{G}$ is embeddable in a subgroup of $\mathbf{T}$. □

Theorem 5.3.2 is proved. □

The main result of this section implies the existence of semigroups with potentially $\mathcal{L}$- and potentially $\mathcal{R}$-embeddable subsets with the property that these subsets are never simultaneously $\mathcal{L}$ and $\mathcal{R}$ related in any embedding semigroup. Such an

example would be found if we could find an extension of rank 3 of a partial group that is not embeddable in a group however such an example is unnecessarily complicated. A much simpler example is found by modifying the second example of the previous section.

EXAMPLE 5.3.5 *To the multiplication table for* $\mathbf{S}$ *in the Example 5.2.8 above, add two elements* d_1, d_2 *with the multiplication* $d_i x = y$ *whenever* $a_i x = y$; $x d_i = y$ *whenever* $x a_i = y$; *and all other products not already defined take the value* 0. *Let the resulting 3-nilpotent semigroup be denoted by* $\mathbf{U}$.

Then the subsets $\{d_1, d_2\}$ *and* $\{a_1, a_2\}$ *of* U *are* $\mathcal{R}^*$ *and* $\mathcal{L}^*$ *classes of* $\mathbf{U}$ *respectively but* $[\{d_1, d_2\}, \{a_1, a_2\}]$ *is not potentially* $\mathcal{R} - \mathcal{L}$*-embeddable.*

Proof: Since $\{a_1, a_2\}$ is an $\mathcal{H}^*$-class of $\mathbf{S}$, then $\{a_1, a_2\}$ and $\{d_1, d_2\}$ lie within $\mathcal{H}^*$-classes of $\mathbf{U}$ (in fact they lie within the same $\mathcal{H}^*$-class) and so certainly they lie in $\mathcal{L}^*$ and $\mathcal{R}^*$-classes respectively.

Now let $\mathbf{T}$ be any semigroup in which $\mathbf{U}$ can be embedded so that a_1 and a_2 are $\mathcal{L}^{\mathbf{T}}$-related. So there is an $x \in T$ such that $x a_1 = a_2$. Therefore,

$$(x b_1) d_1 = x(b_1 d_1) = x c_1 = x a_1 b_1 = a_2 b_1 = c_2 = b_2 d_1.$$

However

$$(x b_1) d_2 = x(b_1 d_2) = x c_3 = x a_1 b_2 = a_2 b_2 = c_3 \neq c_2 = b_2 d_2.$$

So therefore d_1 and d_2 are not $\mathcal{R}^*$-related in $\mathbf{T}$. Hence $[\{d_1, d_2\}, \{a_1, a_2\}]$ is not potentially $\mathcal{R} - \mathcal{L}$-embeddable. □

Let $\mathcal{D}^*$ be defined as $\mathcal{L}^* \vee \mathcal{R}^*$. As a final example in this theme we use a result of [25] to construct an example of $\mathcal{D}^*$-classes of finite semigroups that are not potentially $\mathcal{D}$-embeddable (or $\mathcal{J}$-embeddable) within the class of finite semigroups (recall that every semigroup is potentially $\mathcal{D}$ and $\mathcal{J}$-embeddable in a (possibly infinite) semigroup and that on a finite semigroup, the relations $\mathcal{D}$ and $\mathcal{J}$ coincide; see [10])

EXAMPLE 5.3.6 *Let* $\mathbf{D}$ *be defined by the following 3-nilpotent semigroup:*

$\cdot$	0	a	b	c
0	0	0	0	0
a	0	c	0	0
b	0	c	0	0
c	0	0	0	0

Then the set $\{a, b, c\}$ is a $\mathcal{D}^*$-class of $\mathbf{D}$ but is not potentially $\mathcal{D}$-embeddable (or potentially $\mathcal{J}$-embeddable) in a finite semigroup.

Proof: The $\mathcal{L}^*$ classes of $\mathbf{D}$ are $\{a, b\}$, $\{c\}$, $\{0\}$ and the $\mathcal{R}^*$ classes of $\mathbf{D}$ are $\{a\}$, $\{b, c\}$ and $\{0\}$. Hence $\{a, b, c\}$ is a $\mathcal{D}^*$-class. However if $\{a, b, c\}$ is $\mathcal{D}$-embeddable in a finite semigroup, then it is $\mathcal{D}$-embeddable in a finite 0-simple semigroup. In a finite 0-simple semigroup we have $xyz = 0 \Leftrightarrow xy = 0$ or $yz = 0$ (this property is called *categorical at 0*), however in $\mathbf{D}$ we have $aaa = 0$ with $aa \neq 0$. (This is a direct application of Theorem 2.5 of [25] which states that a 3-nilpotent semigroup is embeddable in a completely 0-simple semigroup if and only if it is categorical at 0.) Hence $\mathbf{D}$ is not embeddable in a finite 0-simple semigroup, and therefore $\{a, b, c\}$ is not potentially $\mathcal{D}$ or $\mathcal{J}$ embeddable within the class of finite semigroups. □

Note that Fountain (Example 2.2 in [20]) has found an 8 element example with $\mathcal{D}^*$-related idempotents e and f satisfying $e > f$ (recall that for idempotents e, f, we define $e < f$ to mean $ef = fe = e$). Since $\mathcal{D}$-classes containing idempotents e, f with $e > f$ are infinite (see [10]) these two elements are not potentially $\mathcal{D}$-embeddable in a finite semigroup.

5.4 On the embeddability of semigroup amalgams

In this section we investigate the problem of determining when an amalgam of semigroups can be embedded in a member of some important class of semigroups. Let K be a class of semigroups and let $\{\mathbf{S}_i : i \in I\}$ be a set of (finite) members of

K indexed by the (finite) set I such that for some semigroup (necessarily finite) $\mathbf{U}$ there are injective homomorphisms $\phi_i : \mathbf{U} \to \mathbf{S}_i$. This collection of semigroups and mappings is called a (finite) K *amalgam* and is denoted by $[\{\mathbf{S}_i : i \in I\}; \mathbf{U}; \{\phi_i : i \in I\}]$ or more briefly $[\mathbf{S}_i; \mathbf{U}; \phi_i]$ or even simply $[\mathbf{S}_i; \mathbf{U}]$. Less formally, a K amalgam may be viewed as a collection of semigroups from K (the $\mathbf{S}_i$) each sharing a common subsemigroup from K (the semigroup $\mathbf{U}$). The semigroup $\mathbf{U}$ is known as the *core* of the amalgam. In these definitions we have not used any specific facts concerning semigroups and indeed we could replace the word "semigroup" in the above by any class of algebraic structures of some fixed type. In several cases we will translate results found for semigroups into related results in ring theory.

An *embedding* of a K amalgam $[\mathbf{S}_i; \mathbf{U}; \phi_i]$ is a set of injective homomorphisms $\{\nu_i : i \in I\}$ with $\nu_i : \mathbf{S}_i \to \mathbf{T}$ for some semigroup $\mathbf{T}$ so that for $s \in \mathbf{S}_i$ and $t \in \mathbf{S}_j$, $\nu_i(s) = \nu_j(t)$ if and only if $i = j$ and $s = t$ or there is a $u \in \mathbf{U}$ such that $\phi_i(u) = s$ and $\phi_j(u) = t$.

The fundamental question to be asked concerning a K amalgam is the following:

QUESTION 5.4.1 *Given a finite K amalgam $\mathbf{A} = [\mathbf{S}_i\mathbf{U}]$, is $\mathbf{A}$ embeddable in a member of K?*

The classes K which we will be primarily concerned with in this section are the class of all semigroups, the class of finite semigroups, the class of all inverse semigroups and the class of finite inverse semigroups. To a lesser degree we will also be interested in similar classes of rings.

For the class of all groups and the class of finite groups, Question 5.4.1 has a remarkably simple solution: the answer is "always" [81]. For semigroups and rings however this is not the case. Consider the following pair of semigroups[2]:

[2] A similar example due to Kimura is presented on page 139 of volume II of [10]

·	0	1	2	3
0	0	0	0	0
1	0	0	0	1
2	0	0	0	1
3	0	1	1	3

·	0	1	2	4
0	0	0	0	0
1	0	0	0	2
2	0	0	0	2
4	0	2	2	4

The two semigroups share a common three element semigroup with zero multiplication and so we may consider them as a semigroup amalgam. However this amalgam is not embeddable in any semigroup $\mathbf{T}$ since in that case we would have

$$3 \cdot (1 \cdot 4) = 3 \cdot 2 = 1 \text{ and } (3 \cdot 1) \cdot 4 = 1 \cdot 4 = 2.$$

(Here we regard $\{0,1,2,3,4\}$ as being a subset of $\mathbf{T}$, and the maps ϕ_1, ϕ_2, ν_1, ν_2 as being the identity maps on their domains.) That is, associativity fails in any groupoid in which the amalgam is embeddable. It is clear that a semigroup amalgam A determines a partial groupoid in a natural way but the example above shows that this is *not necessarily* a partial semigroup in the sense that we do *not necessarily* have $(xy)z = x(yz)$ whenever both sides of this expression are defined.

Question 5.4.1 for rings and semigroups has consequently been the subject of a substantial quantity of work and several books on semigroup theory contain a chapter devoted to it and associated concepts. More generally we may formulate the following decision problems:

PROBLEM 5.4.2 *(i) Given a finite semigroup (ring) amalgam* $\mathbf{A} = [\mathbf{S}_i; \mathbf{U}]$ *of semigroups (rings) from a class* K*, determine if* $\mathbf{A}$ *is embeddable in a semigroup (ring) from* K*.*
(ii) Given a finite semigroup (ring) amalgam $\mathbf{A} = [\mathbf{S}_i; \mathbf{U}]$*, determine if* $\mathbf{A}$ *is embeddable in a semigroup (ring) from a class* K*.*

We will call problems 5.4.2 (i) and (ii) respectively (within the class of semigroups or rings) *the strong decision problem for amalgam embeddability in* K and *the weak*

decision problem for amalgam embeddability in K respectively. Note that a negative answer to the strong decision problem for a class K implies a negative answer to the weak decision problem for K. In terms of decidability and undecidability these problems will coincide for some classes. For example if K is a variety then any semigroup (ring) amalgam containing a subsemigroup (subring) not from K (that is, not satisfying one of the defining identities of K) clearly is not embeddable in a member of K. This is similar to trying to embed a non-associative groupoid (non-associative ring) in a semigroup (ring). Recent results of Kublanovsky and Sapir [45] can be used to show that strong and weak decision problems for embeddability of semigroup (ring) amalgams in the class of finite semigroups (rings) are undecidable (see Theorem 5.4.10). The main result we prove in this section is the following theorem.

THEOREM 5.4.3 *There is no algorithm to decide when given an arbitrary finite semigroup (ring) amalgam $\mathbf{A} = [\mathbf{S}_i; \mathbf{U}]$ whether $\mathbf{A}$ is embeddable in a semigroup (ring). That is, the strong and weak decision problems for embeddability of amalgams in the class of semigroups (rings) and in the class of finite semigroups (finite rings) are undecidable.*

In particular Problems 5.4.2 part (i) and part (ii) are undecidable. We note that there are several important classes for which the corresponding problems have a very different solution. We have seen that any finite group amalgam can be embedded in a finite group. Similarly any finite amalgam of inverse semigroups can be embedded in an inverse semigroup (see [29]), however this is not necessarily finite (see page 309 of [29] for an example, due to C. J. Ash, of a finite inverse semigroup amalgam not embeddable in a finite semigroup). Interestingly, we will show that the *weak* decision problem for inverse semigroups and finite inverse semigroups is undecidable. The class of subsemigroups of inverse semigroups has a decidable membership problem (see [10] for a description due to B. Schein). However if K is a class closed under taking subsemigroups (or subrings respectively) with undecidable

membership problem then the weak decision problem for amalgam embeddability in K is undecidable since individual semigroups (rings) can be considered, trivially, as amalgams by themselves ($[\mathbf{S}; \mathbf{S}]$ in our notation). A good example of a class closed under the taking of subsemigroups that has undecidable membership problem is the class of subsemigroups of completely 0-simple semigroups which was shown to have this property by Kublanovsky (see [25]). Naturally, this argument does not apply to the strong decision problem for amalgam embeddability.

A generalization of amalgam embeddability is *weak amalgam embeddability* (see [29]). If $\mathbf{A}$ is a semigroup (ring) amalgam $[\mathbf{S}_i; \mathbf{U}; \phi_i]$ then we will say $\mathbf{A}$ is *weakly embeddable* in a semigroup (ring) $\mathbf{T}$ if for each i there are injective homomorphisms $\nu_i : \mathbf{S}_i \to \mathbf{T}$ such that for every $u \in \mathbf{U}$, $\phi_i(u) = s$ and $\phi_j(u) = t$ imply $\nu_i(s) = \nu_j(t)$. So any embedding of an amalgam is a weak embedding but not every weak embedding is an embedding. We can replace "embeddable" with "weakly embeddable" in Problem 5.4.2 (i) and (ii) and call the respective decision problems *the strong decision problem for weak amalgam embeddability in* K and *the weak decision problem for weak amalgam embeddability in* K. It is conceivable that a class K has an undecidable (strong or weak) decision problem for amalgam embeddability but a decidable (strong or weak) decision problem for weak amalgam embeddability (or vice versa). We will show that this is not the case for the class of all rings (semigroups) and the class of finite rings (semigroups).

THEOREM 5.4.4 *The strong and weak decision problems for weak embeddability of ring (semigroup) amalgams in the class of all semigroups (rings) and in the class of finite semigroups (finite rings) are undecidable.*

Let $\mathbf{S}$ be a semigroup and $\mathbb{Z}_2$ be the field of two elements, $\{0, 1\}$. Then the universe of the semigroup ring $\mathbb{Z}_2[\mathbf{S}]$ is the set of all functions $f : \mathbf{S} \to \{0, 1\}$ which map only finitely many elements of $\mathbf{S}$ to 1. The addition on $\mathbb{Z}_2[\mathbf{S}]$ is pointwise and the multiplication is defined by $fg(s) = \sum_{s_i s_j = s} f(s_i)g(s_j)$. There is a natural

embedding of every semigroup $\mathbf{S}$ into the multiplicative semigroup of the semigroup ring $\mathbb{Z}_2[\mathbf{S}]$ which sends an element s to the function f_s defined by $f_s(t) = 1$ if $s = t$ and 0 otherwise. Also if $\mathbf{S}$ is a subsemigroup of a semigroup $\mathbf{T}$ then by considering those elements of $\mathbb{Z}_2[\mathbf{S}]$ which are functions sending all elements $t \in \mathbf{T} \backslash \mathbf{S}$ to 0 we have that the semigroup ring $\mathbb{Z}_2[\mathbf{S}]$ is a subring of $\mathbb{Z}_2[\mathbf{T}]$. These facts enable one to translate many semigroup embedding problems into ring embedding problems. This will also be true of amalgam embeddability.

Given a semigroup amalgam $\mathbf{A} = [\mathbf{S}_i; \mathbf{U}]$ we can construct the ring amalgam $\mathbb{Z}_2[\mathbf{A}] = [\mathbb{Z}_2[\mathbf{S}_i]; \mathbb{Z}_2[\mathbf{U}]]$. The amalgam $\mathbf{A}$ can be embedded into the multiplicative semigroup amalgam of $\mathbb{Z}_2[\mathbf{A}]$ as a "sub-amalgam" in the natural way. If $\mathbf{A}$ is (weakly) embeddable in $\mathbf{T}$ then $\mathbb{Z}_2[\mathbf{A}]$ is (weakly) embeddable in $\mathbb{Z}_2[\mathbf{T}]$ (which is finite if and only if $\mathbf{T}$ is). Furthermore, if $\mathbb{Z}_2[\mathbf{A}]$ is (weakly) embeddable in a ring or finite ring $\mathbf{R}$ then the amalgam $\mathbf{A}$ (which is a "sub-amalgam" of the multiplicative semigroup amalgam of the ring amalgam $\mathbb{Z}_2[\mathbf{A}]$) is (weakly) embeddable in the multiplicative semigroup of $\mathbf{R}$. Thus it will suffice to prove Theorems 5.4.3 and 5.4.4 in the case of semigroups. This will be done in the style of the previous results in this chapter. However the role of partial groups will be replaced by the slightly more specific symmetric partial groups.

A *symmetric partial group* (see [45]) is a partial group $\mathbf{G}$ with the property that for every $g \in \mathbf{G}$ there is a unique $g' \in \mathbf{G}'$ such that $gg' = g'g = 1$. For any finite partial group we may construct a *symmetric extension* $\mathbf{G}'$ of $\mathbf{G}$ which is a symmetric partial group containing $\mathbf{G}$ such that for every $g \in \mathbf{G}'$, either g or g' is contained in the partial group $\mathbf{G}$. This condition ensures that there are only finitely many possible symmetric extensions and they may be effectively listed. It is also clear that if $\mathbf{G}$ is embeddable in a group then there is a symmetric extension of $\mathbf{G}$ that is embeddable in a group, since every group may be considered as a symmetric partial group (where the "partial" operation is defined everywhere). Thus the problem of determining whether a finite symmetric partial group is embeddable in a group or

in a finite group is also undecidable.

For all arguments to follow we will take $\mathbf{G_4}$ to be an extension of rank four of a finite symmetric partial group $\mathbf{G_1}$ and G_i will be used to denote the set $\cup_{j=0}^{i} G^j$ for $i \leq 4$.

DEFINITION 5.4.5 *For any group* $\mathbf{G}$, *we will let* $\mathbf{B_n(G)}$ *denote the Brandt semigroup* $\mathcal{M}[\mathbf{G}, n, n, I]$ *where* I *is an* $n \times n$ *identity matrix over* $G \cup \{0\}$.

Despite the apparently simple structure of the Brandt semigroups, in [25] it is shown that the set of finite subsemigroups of Brandt semigroups is not recursive. It is well known that all Brandt semigroups are inverse semigroups.

We now define a finite semigroup $\mathbf{S(G_1, G_4)}$ corresponding to any finite extension $\mathbf{G_4}$ of rank four of a symmetric partial group $\mathbf{G_1}$.

DEFINITION 5.4.6 *Let* $\mathbf{G_4}$ *be an extension of rank four of a symmetric partial group* $\mathbf{G_1}$ *with* $G_0, G_1, \ldots, G_4$ *defined as before. Then we construct the semigroup* $\mathbf{S(G_1, G_4)}$ *on the set*

$$\{(i, g, j) : 0 < i \leq j \leq 5, g \in G_{j-i}\} \cup \{0\}$$

with the multiplication $(i, g, j) \cdot (k, h, l) = (i, gh, l)$ *if* $j = k$ *and* gh *is the product of* g *with* h *in* $\mathbf{G_4}$ *and* 0 *otherwise.*

As in the constructions presented in the previous two sections, it is not difficult to verify that this is indeed a semigroup. Associativity holds essentially because we required it to be so in our definition of an extension of rank k. If $\mathbf{G_4}$ is embeddable in a group $\mathbf{H}$ then $\mathbf{S(G_1, G_4)}$ can be viewed as a subsemigroup of "the upper half" of the Brandt semigroup $\mathbf{B_5(H)}$ over $\mathbf{H}$.

Let $\langle \mathbf{1} \rangle$ be the one element group. Now the intersection of the universe of $\mathbf{B_5}(\langle \mathbf{1} \rangle)$ with the universe of $\mathbf{S(G_1, G_4)}$ consists of those elements of $\mathbf{B_5}(\langle \mathbf{1} \rangle)$ of the form $(i, 1, j)$ where $i \leq j$. Furthermore the restriction of the operations of both $\mathbf{B_5}(\langle \mathbf{1} \rangle)$

and $\mathbf{S}(\mathbf{G_1}, \mathbf{G_4})$ to this set coincide and form a subsemigroup. We will denote this subsemigroup by $\mathbf{S}(\langle \mathbf{1} \rangle, \langle \mathbf{1} \rangle)$ which is consistent with our previous definition, since $\langle \mathbf{1} \rangle$ can be considered as an extension of rank four of itself. We can now construct the following amalgam.

DEFINITION 5.4.7 *For a finite extension* $\mathbf{G_4}$ *of rank four of a symmetric partial group* $\mathbf{G_1}$ *define an associated semigroup amalgam* $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ *by*

$$\mathbf{A}(\mathbf{G_1}, \mathbf{G_4}) = [\mathbf{S}(\mathbf{G_1}, \mathbf{G_4}), \mathbf{B_5}(\langle \mathbf{1} \rangle); \mathbf{S}(\langle \mathbf{1} \rangle, \langle \mathbf{1} \rangle)]$$

The following tables representing $\mathbf{S}(\mathbf{G_1}, \mathbf{G_4})$ and $\mathbf{B_5}(\langle \mathbf{1} \rangle)$ respectively may help to visualize the amalgam we have constructed (here $(i, G, j) = \{(i, g, j) : g \in G\}$):

$(1, G_0, 1)$	$(1, G_1, 2)$	$(1, G_2, 3)$	$(1, G_3, 4)$	$(1, G_4, 5)$
	$(2, G_0, 2)$	$(2, G_1, 3)$	$(2, G_2, 4)$	$(2, G_3, 5)$
		$(3, G_0, 3)$	$(3, G_1, 4)$	$(3, G_2, 5)$
			$(4, G_0, 4)$	$(4, G_1, 5)$
				$(5, G_0, 5)$

$(1, G_0, 1)$	$(1, G_0, 2)$	$(1, G_0, 3)$	$(1, G_0, 4)$	$(1, G_0, 5)$
$(2, G_0, 1)$	$(2, G_0, 2)$	$(2, G_0, 3)$	$(2, G_0, 4)$	$(2, G_0, 5)$
$(3, G_0, 1)$	$(3, G_0, 2)$	$(3, G_0, 3)$	$(3, G_0, 4)$	$(3, G_0, 5)$
$(4, G_0, 1)$	$(4, G_0, 2)$	$(4, G_0, 3)$	$(4, G_0, 4)$	$(4, G_0, 5)$
$(5, G_0, 1)$	$(5, G_0, 2)$	$(5, G_0, 3)$	$(5, G_0, 4)$	$(5, G_0, 5)$

Theorems 5.4.3 and 5.4.4 now follow from the following theorem.

THEOREM 5.4.8 *Let* $\mathbf{G_1}$ *be a finite symmetric partial group. The following are equivalent:*

(i) $\mathbf{G_1}$ *is embeddable in a group (finite group);*

(ii) There is an extension $\mathbf{G_4}$ *of rank four of* $\mathbf{G_1}$ *that is embeddable in a group (finite group);*

(iii) There is an extension $\mathbf{G_4}$ *of rank four of* $\mathbf{G_1}$ *such that* $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ *is embeddable in a Brandt semigroup (finite Brandt semigroup);*

(iv) There is an extension $\mathbf{G_4}$ *of rank four of* $\mathbf{G_1}$ *such that* $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ *is embeddable in a semigroup (finite semigroup);*

(v) There is an extension $\mathbf{G_4}$ *of rank four of* $\mathbf{G_1}$ *such that* $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ *is weakly embeddable in a Brandt semigroup (finite Brandt semigroup);*

(vi) There is an extension $\mathbf{G_4}$ *of rank four of* $\mathbf{G_1}$ *such that* $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ *is weakly embeddable in a semigroup (finite semigroup).*

Proof: That (i)⇒(ii) follows from comments following the definition of an extension of rank k of a partial group.

$(ii) \Rightarrow (iii)$: Say $\mathbf{G_4}$ is embeddable in a group $\mathbf{H}$. Then it is easily verified that $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ is embedded in the Brandt semigroup $\mathbf{B_5}(\mathbf{H})$ by the identity maps: $\nu_1 : \mathbf{S}(\mathbf{G_1}, \mathbf{G_4}) \to \mathbf{B_5}(\mathbf{H})$ and $\nu_2 : \mathbf{B_5}(\langle \mathbf{1} \rangle) \to \mathbf{B_5}(\mathbf{H})$ that take an element from their respective domains and assign to it the element with the same name in $\mathbf{B_5}(\mathbf{H})$. Note that $\mathbf{B_5}(\mathbf{H})$ is an inverse semigroup that is finite if and only if $\mathbf{H}$ is a finite group.

(iii)⇒(iv) and (v)⇒(vi): Trivial.

(iii)⇒(v) and (iv)⇒(vi): This follows since every embedding of an amalgam is a weak embedding of that amalgam.

(vi)⇒(i): Say the amalgam $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ is weakly embeddable in a semigroup $\mathbf{T}$ (a finite semigroup $\mathbf{T}$) via the injective homomorphisms $\nu_1 : \mathbf{S}(\mathbf{G_1}, \mathbf{G_4}) \to \mathbf{T}$ and $\nu_2 : \mathbf{B_5}(\langle \mathbf{1} \rangle) \to \mathbf{T}$.

For any element $g \in \mathbf{G_1}$ consider the element $\nu_1(2, g, 3)$ of $\mathbf{T}$. Now since $\mathbf{G_1}$ is symmetric, there is an element g' such that $gg' = g'g = 1$ in $\mathbf{G_1}$ (and of course in $\mathbf{G_4}$ since $\mathbf{G_1}$ is contained within $\mathbf{G_4}$). So $\nu_1(2, g, 3)[\nu_1(3, g', 4)] = \nu_1((2, g, 3)(3, g', 4)) =$

$\nu_1((2,1,4))$ and

$$\begin{aligned}
\nu_1((2,1,4))[\nu_2((4,1,2))\nu_1((2,g,3))] =& [\nu_1((2,1,4))\nu_2((4,1,2))]\nu_1((2,g,3)) \\
& \text{(by associativity)} \\
=& [\nu_2((2,1,4))\nu_2((4,1,2))]\nu_1((2,g,3)) \\
& \text{(since } \nu_1((2,1,4)) = \nu_2((2,1,4))\text{)} \\
=& [\nu_2((2,1,4)(4,1,2))]\nu_1((2,g,3)) \\
=& [\nu_2((2,1,2))]\nu_1((2,g,3)) \\
=& [\nu_1((2,1,2))]\nu_1((2,g,3)) \\
& \text{(since } \nu_1((2,1,2)) = \nu_2((2,1,2))\text{)} \\
=& \nu_1((2,g,3)).
\end{aligned}$$

Note that we do not know what the product $\nu_2((4,1,2))\nu_1((2,g,3))$ from the first line actually is in $\mathbf{T}$, only that it does exist. Therefore the set $H_{2,3} = \{\nu_1((2,g,3)) : g \in \mathbf{G_1}\}$ is $\mathcal{R}$-related to $\nu_1((2,1,4))$. In particular $H_{2,3}$ lies within an $\mathcal{R}$-class of $\mathbf{T}$.

Also $[\nu_1((1,g',2))]\nu_1((2,g,3)) = \nu_1((1,g',2)(2,g,3)) = \nu_1((1,1,3))$ and

$$\begin{aligned}
[\nu_1((2,g,3))\nu_2((3,1,1))]\nu_1((1,1,3)) =& \nu_1((2,g,3))[\nu_2((3,1,1))\nu_1((1,1,3))] \\
=& \nu_1((2,g,3))[\nu_2((3,1,1))\nu_2((1,1,3))] \\
=& \nu_1((2,g,3))[\nu_2((3,1,1)(1,1,3))] \\
=& \nu_1((2,g,3))[\nu_2((3,1,3))] \\
=& \nu_1((2,g,3))[\nu_1((3,1,3))] \\
=& \nu_1((2,g,3)).
\end{aligned}$$

Thus $H_{2,3}$ is within an $\mathcal{L}$-class of $\mathbf{T}$. In particular since $H_{2,3}$ is both $\mathcal{L}$- and $\mathcal{R}$-related in $\mathbf{T}$, it lies within an $\mathcal{H}$-class of $\mathbf{T}$.

Now for each $g \in \mathbf{G_1}$ we can consider the element $\nu_1((3,g,4))$ of $\mathbf{T}$. Replacing every expression of the form (i,h,j) in the above arguments by $(i+1,h,j+1)$ we

obtain the analogous result that the set $H_{3,4} = \{(3, g, 4) : g \in \mathbf{G_1}\}$ is also contained in an $\mathcal{H}$-class of $\mathbf{T}$.

Consider the extension $\mathbf{G_2}$ of rank 2 of $\mathbf{G_1}$ consisting of the elements of the set G_2 with the partial operation $f \cdot g = h$ if and only if either f or g is contained in the set G_1 and $fg = h$ in $\mathbf{G_4}$. We can construct the associated split system $(\{a\} \times G_1 \times \{b\}, \{b\} \times G_1 \times \{c\}, \{a\} \times G_2 \times \{c\})$ and a corresponding embedding (j, k, l) into $\mathbf{T}$ defined by

$$j((a, g, b)) = \nu_1((2, g, 3)),\ k((b, g, c)) = \nu_1((3, g, 4)),\ j((a, g, c)) = \nu_1((2, h, 4))$$

where $x \in \{j, k, l\}$, $y \in \{a, b\}$, $z \in \{b, c\}$, $y \neq z$, and g is contained in G_1 and h is contained in G_2. It is clear that these maps are injective and constitute an embedding of $(\{a\} \times G_1 \times \{b\}, \{b\} \times G_1 \times \{c\}, \{a\} \times G_2 \times \{c\})$ since $j((a, g, b))k((b, h, c)) = \nu_1((2, g, 3))\nu_1((3, h, 4)) = \nu_1((2, g, 3)(3, h, 4)) = \nu_1((2, gh, 4)) = l((a, gh, c))$. Furthermore since the images of j and k are the sets $H_{2,3}$ and $H_{3,4}$ respectively and these lie within $\mathcal{H}$-classes of $\mathbf{T}$ we may apply Lemma 5.1.6 to show that $\mathbf{G_1}$ is embeddable in a group. The Theorem is proved. □

The two semigroups $\mathbf{B_5}(\langle \mathbf{1} \rangle)$ and $\mathbf{S}(\langle \mathbf{1} \rangle, \langle \mathbf{1} \rangle)$ involved in the amalgams used for this proof are fixed throughout. Furthermore since Brandt semigroups are inverse semigroups we have actually proved the following result.

COROLLARY 5.4.9 *There is no algorithm that determines when given a finite semigroup amalgam* $\mathbf{A} = [\mathbf{S_1}, \mathbf{S_2}; \mathbf{U}]$ *with* $|S_2| \leq 26$, $|U| \leq 16$, *whether* $\mathbf{A}$ *is embeddable (or weakly embeddable) in any of the following: a semigroup; a finite semigroup; an inverse semigroup; a finite inverse semigroup.*

So the weak decision problem for amalgam embeddability in the class of inverse semigroups and finite inverse semigroups is undecidable.

In the case of embedding (weak or otherwise) a semigroup amalgam in a finite semigroup (or in a finite inverse semigroup) we may improve the bounds in this theorem as follows.

THEOREM 5.4.10 *There is no algorithm that determines when given a finite semigroup amalgam* $\mathbf{A} = [\mathbf{S}_1, \mathbf{S}_2; \mathbf{U}]$ *with* $|S_2| \leq 7$, $|U| \leq 5$, *whether* $\mathbf{A}$ *is embeddable in a finite semigroup or a finite inverse semigroup.*

Proof: This essentially follows from the main result in [45]. For any extension $\mathbf{G_3}$ of rank 3 of a partial group $\mathbf{G_1}$ (with G_i for $i \leq 3$ defined as before) we may construct a semigroup $\mathbf{S}(\mathbf{G_1}, \mathbf{G_3})$ in the following way: the universe of $\mathbf{S}(\mathbf{G_1}, \mathbf{G_3})$ is the set $\{(i, g, j) : 0 < i < j \leq 4, g \in G_{j-i}\}$ and the multiplication is defined in the same way as that for $\mathbf{S}(\mathbf{G_1}, \mathbf{G_4})$ in Definition 5.4.6 (this semigroup first appeared in [25]). In [45], Kublanovsky and Sapir show that for a symmetric partial group $\mathbf{G_1}$ one can find an extension $\mathbf{G_3}$ of rank three of $\mathbf{G_1}$ embeddable in a finite group, if and only if one can find a finite semigroup $\mathbf{T}$ containing $\mathbf{S}(\mathbf{G_1}, \mathbf{G_3})$ with elements $x, y \in T^1$ such that $x \cdot (1, 1, 4) \cdot y = (2, 1, 3)$ in $\mathbf{T}$. With this in mind, we can construct an amalgam consisting of $\mathbf{S}(\mathbf{G_1}, \mathbf{G_3})$ along with a semigroup that *enforces* this condition in any embedding semigroup. This second semigroup, $\mathbf{S_2}$, can be taken as the set

$$\{(2, 1, 1), (4, 1, 3), (2, 1, 3), (1, 1, 4), (2, 1, 4), (1, 1, 3), 0\}$$

with multiplication as within a Brandt semigroup. The set

$$U = \{(2, 1, 3), (1, 1, 4), (2, 1, 4), (1, 1, 3), 0\}$$

is common to both $\mathbf{S_2}$ and $\mathbf{S}(\mathbf{G_1}, \mathbf{G_3})$ and furthermore the restriction of the operations of these semigroups to U coincide and forms a subsemigroup of both which we will call $\mathbf{U}$. It is now easily verified that the following constitutes a semigroup amalgam:

$$\mathbf{A}'[\mathbf{G_1}, \mathbf{G_3}] = [\mathbf{S}(\mathbf{G_1}, \mathbf{G_3}), \mathbf{S_2}; \mathbf{U}].$$

Furthermore if $\mathbf{G_3}$ is embeddable in a finite group $\mathbf{H}$, then this amalgam is embeddable in $\mathbf{B_4}(\mathbf{H})$ in the obvious way (analogous to the embedding of $\mathbf{A}(\mathbf{G_1}, \mathbf{G_4})$ into $\mathbf{B_5}(\mathbf{H})$ in the proof of Theorem 5.4.8). On the other hand, if $\mathbf{A}'(\mathbf{G_1}, \mathbf{G_3})$ is

embeddable into a finite semigroup $\mathbf{T}$ by the maps ν_1 and ν_2 then we have

$$\begin{aligned}\nu_2((2,1,1))\nu_1((1,1,4))\nu_2((4,1,3)) &= \nu_2((2,1,1))\nu_2((1,1,4))\nu_2((4,1,3)) \\ &= \nu_2((2,1,3)) \\ &= \nu_1((2,1,3))\end{aligned}$$

and therefore $\mathbf{G_1}$ is embeddable in a group. □

Note that there is a subsemigroup $\mathbf{U}'$ of $\mathbf{U}$ on the three element set

$$\{(2,1,3),(1,1,4),0\}$$

If we replace $\mathbf{U}$ by $\mathbf{U}'$ throughout the proof of Theorem 5.4.10, all arguments remain valid except the natural embedding of $\mathbf{A}'(\mathbf{G_1},\mathbf{G_3})$ into $\mathbf{B_4}(\mathbf{H})$ is now only a *weak* amalgam embedding since $\nu_1((2,1,4)) = \nu_2((2,1,4))$ though $(2,1,4) \notin \mathbf{U}'$. Thus we have proved

THEOREM 5.4.11 *There is no algorithm that determines when given a semigroup amalgam* $\mathbf{A} = [\mathbf{S}_1, \mathbf{S}_2; \mathbf{U}]$ *with* $|S_2| = 7$, $|U| = 3$, *whether* $\mathbf{A}$ *is weakly embeddable in a finite semigroup.*

In [77], Sapir proves the undecidability of the strong decision problem for amalgam embeddability in the class of finite semigroups using an almost identical structure to that we use to prove Theorem 5.4.10 above however the bounds for $|S_2|$ and $|U|$ are 17 and 7 respectively.

Fundamental to the proof of Kublanovsky and Sapir's result [45] is the fact that finite semigroups consisting of only one non zero $\mathcal{J}$-class have a particularly well defined structure: they are completely 0-simple and by a well known theorem of Rees, isomorphic to a Rees Matrix semigroup with zero over a group (see [10] or [29] for details). The completely 0-simple structure is not available in the general case of embedding in a $\mathcal{J}$-class of an arbitrary semigroup (indeed *any* finite semigroup can

be embedded in an infinite semigroup with a single $\mathcal{J}$-class which is *not* completely 0-simple), and this is why the proofs of Theorem 5.4.10 and Theorem 5.4.11 only apply for embedding amalgams in the class of finite semigroups.

NOTE 5.4.12 *Theorems 5.4.9, 5.4.10 and 5.4.11 have ring analogues. To obtain these we can replace "semigroup" with "ring" and any numbers n appearing in the theorems by* 2^n.

This is because if $\mathbf{S}$ is a finite semigroup with n elements then the semigroup ring $\mathbb{Z}_2[\mathbf{S}]$ has 2^n elements.

Necessary and sufficient conditions for the embeddability of a semigroup amalgam into a semigroup have been found by Howie [27]. We will describe this characterisation since by Theorem 5.4.3 the conditions involved must not be algorithmically verifiable.

Let $\mathbf{A} = [\{\mathbf{S}_i : i \in I\}; \mathbf{U}; \{\phi_i : i \in I\}]$ be a semigroup amalgam. We will assume that the sets S_i are pairwise disjoint (here, as usual, S_i denotes the universe of $\mathbf{S}_i$). The *free product*, $\Pi^*\mathbf{S}_i$, is the semigroup generated by the set $X_A = \cup S_i$ with the Cayley tables of the $\mathbf{S}_i$ determining the relations R_A. That is, $\Pi^*\mathbf{S}_i$ is the semigroup $\langle X_A; R_A \rangle$. We may define a congruence θ on $\Pi^*\mathbf{S}_i$ as the congruence generated by $\{(\phi_i(u), \phi_j(u)) : i, j \in I,\ u \in U\}$. The free product of the amalgam $\mathbf{A}$ is the semigroup $\Pi^*_U\mathbf{S}_i = \langle X_A; R_A \rangle / \theta$. For each $i \in I$ there are homomorphisms ν_i from each $\mathbf{S}_i$ into $\Pi^*_U\mathbf{S}_i$ defined by $\nu_i(s) = s$. If these maps constitute an embedding of the amalgam $\mathbf{A}$ then it is said that $\mathbf{A}$ is naturally embedded in its free product.

THEOREM 5.4.13 *[27] The amalgam* $\mathbf{A}$ *is embeddable in a semigroup if and only if it is naturally embedded in its free product.*

Let X'_A be the set

$$U \cup (X_A \backslash \{\phi_i(u) : u \in U,\ i \in I\})$$

and R'_A determined by the set of Cayley tables of the $\mathbf{S}_i$ with every occurrence of an element of the form $\phi_i(u)$ replaced by the element u. We have

$$\Pi^*_U \mathbf{S}_i = \langle X_A; R_A \rangle / \theta \;\cong\; \langle X'_A; R'_A \rangle.$$

The previous theorem can now be restated as

THEOREM 5.4.14 *The amalgam* $\mathbf{A}$ *is embeddable in a semigroup if and only if the elements* X'_A *are distinct in* $\langle X'_A; R'_A \rangle$.

Thus we may reformulate Theorem 5.4.3 for semigroups as

COROLLARY 5.4.15 *There is no algorithm that will solve the following decision problem: given a finite semigroup amalgam* $\mathbf{A}$*, determine whether two generators* $x, y \in X_A$ *represent different elements of the semigroup* $\langle X'_A; R'_A \rangle$.

We finish with some questions.

QUESTION 5.4.16 *(i) What are the minimal pairs* $(|S_1|, |U|)$ *for which Theorems 5.4.9, 5.4.10 or 5.4.11 (or their ring analogues) are true and are these minimal pairs the same?*
(ii) Are there classes for which the decision problem for amalgam embeddability is decidable (or undecidable) and the decision problem for weak amalgam embeddability is undecidable (or decidable respectively)?
(iii) Are there varieties $\mathcal{V}$ *for which the (strong or weak) decision problem for amalgam embeddability or weak amalgam embeddability is decidable (or undecidable) but the opposite is true for the finite trace of* $\mathcal{V}$ *(that is, the finite members of* $\mathcal{V}$*)?*

Regarding the first of these questions we note that in [25] it is shown that any semigroup amalgam with a two element core is embeddable in a semigroup. The last question seems of particular interest when $\mathcal{V}$ is the class of inverse semigroups (which form a variety in the signature $\{\cdot, ^{-1}\}$) since it is known that every inverse semigroup amalgam is embeddable in an inverse semigroup, but also that not every

such finite amalgam is embeddable in a *finite* inverse semigroup (see [29]). We note however that one of the main results of [60] shows that there is an algorithm that determines, given a finite semigroup amalgam $\mathbf{A}$ with inverse semigroup core, whether $\mathbf{A}$ is embeddable in a finite semigroup, though the embedding semigroup is not inverse.

Appendix A

Ten small WFB semigroups that generate varieties with uncountably many subvarieties.

Here we list the Cayley tables some seven element WFB semigroups that generate varieties with uncountably many subvarieties. That each of these semigroups is WFB follows from results of [74] (see Theorem 1.1.2). The first seven monoids have index three and therefore by Theorem 4.1.2 generate varieties with uncountably many subvarieties if and only if they do not satisfy $xyx \approx yxx$ or $xyx \approx xxy$. It is a routine matter to verify that these identities are not satisfied by any of the semigroups below. The eighth example is isomorphic to $S(\{aba\})$ and therefore has the desired property by Theorem 4.1.6. The final two examples are isomorphic to the seven element semigroup described in Example 4.2.9 and a corresponding example constructed from $\mathbf{A_2}$. Finite bases of identities have not been established for any but $S(\{aba\})$ (see above) and the first example below. It is possible to show that the closure under deletion of letters of the following set of identities is basis for the

semigroup identities of this semigroup.

$$\{xxx \approx xxxx\}$$

$$\cup \{xu_1xu_2xu_3x \approx xu_1u_2xu_3x,\ xu_1xu_2xu_3x \approx xu_1xu_2u_3x\}$$

$$\cup \{xu_1yu_2xy \approx xu_1yu_2yx,\ xyu_1xu_2y \approx yxu_1xu_2y\}$$

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	0	0	a	0	0
b	0	b	a	b	a	0	0
c	0	c	0	0	c	0	0
d	0	d	0	0	0	e	0
e	0	e	0	0	0	0	0

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	0	0	a	0	0
b	0	b	a	b	0	0	0
c	0	c	0	0	c	0	0
d	0	d	0	0	0	e	0
e	0	e	0	0	0	0	0

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	0	0	0	0	0
b	0	b	0	0	b	0	0
c	0	c	a	0	c	0	0
d	0	d	0	0	0	e	0
e	0	e	0	0	0	0	0

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	0	0	0	0	0
b	0	b	a	b	b	0	0
c	0	c	a	c	c	0	0
d	0	d	0	0	0	e	0
e	0	e	0	0	0	0	0

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	1	c	b	d	e
b	0	b	b	0	0	0	0
c	0	c	c	0	0	0	0
d	0	d	d	0	0	e	0
e	0	e	e	0	0	0	0

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	1	c	b	d	e
b	0	b	b	b	b	0	0
c	0	c	c	c	c	0	0
d	0	d	d	0	0	e	0
e	0	e	e	0	0	0	0

·	1	a	b	c	d	e	f
1	1	a	b	c	d	e	f
a	a	a	a	a	a	a	a
b	b	a	a	a	c	a	a
c	c	c	c	c	c	c	c
d	d	d	d	d	d	d	d
e	e	a	a	a	a	f	a
f	f	a	a	a	a	a	a

·	0	1	a	b	c	d	e
0	0	0	0	0	0	0	0
1	0	1	a	b	c	d	e
a	0	a	0	c	0	e	0
b	0	b	d	0	0	0	0
c	0	c	e	0	0	0	0
d	0	d	0	0	0	0	0
e	0	e	0	0	0	0	0

·	0	a	b	c	d	e	f
0	0	0	0	0	0	0	0
a	0	0	c	0	a	0	0
b	0	d	0	b	0	0	0
c	0	a	0	c	0	0	0
d	0	0	b	0	d	0	0
e	0	0	0	0	0	0	e
f	0	e	0	0	0	e	f

·	0	a	b	c	d	e	f
0	0	0	0	0	0	0	0
a	0	a	b	a	b	0	0
b	0	0	0	a	b	0	0
c	0	c	d	c	d	0	0
d	0	0	0	c	d	0	0
e	0	0	0	0	0	0	e
f	0	e	0	0	0	e	f

Bibliography

[1] E. Artin. "The Gamma function" (English translation by M. Butler), Holt, Rinehart and Winston, 1964.

[2] D. B. Bean, A. Ehrenfeucht, and G. McNulty. Avoidable patterns in strings of symbols, Pacific J. Math. **85** (1979), 261-294.

[3] J. A. Beachy and W. D. Blair. "Abstract Algebra with a Concrete Introduction", Prentice-Hall Inc., 1990.

[4] C. J. Birget, S. Margolis and J. Rhodes. Semigroups whose idempotents form a subsemigroup, Bull. Austral. Math. Soc. **41** (1990), 161-184.

[5] G. Birkhoff. On the structure of abstract algebra, Proc. Camb. Philos. Soc. **28** (1935), 433-454.

[6] A. P. Birjukov. Varieties of idempotent semigroups (Russian), Algebra i Logika **9** (1970), 255-273.

[7] R. C. Buck. "Advanced Calculus", McGraw-Hill Book Company, 3rd ed., 1978.

[8] W. Burnside. "Theory of Groups of Finite Order", Cambridge University Press, 2nd ed., 1911.

[9] S. Burris and H. Sankappanavar. "A Course in Universal Algebra", Grad. Texts in Math. **78**, Springer Verlag, 1981.

[10] A. H. Clifford and G. B. Preston. "The Algebraic Theory of Semigroups," Volumes 1,2, Amer. Math. Soc., Providence, RI, 1961, 1967.

[11] S. Crvenković and D. Delić. A variety with locally solvable but globally unsolvable word problem, Algebra Universalis **35** (1996), 420-424.

[12] S. Crvenković and I. Dolinka. A variety with undecidable equational theory and solvable word problem, Internat. J. Algebra Comput. **8** (1998), 625-630.

[13] D. Delić. From multisorted structures to pseudorecursive varieties, to appear in Trans. Amer. Math. Soc..

[14] C. C. Edmunds. On certain finitely based varieties of semigroups, Semigroup Forum **15** (1977), 21-39.

[15] C. C. Edmunds. Varieties generated by semigroups of order four, Semigroup Forum **21** (1980), 67-81.

[16] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi. "Higher Transcendental Functions," the Bateman Manuscript Project, Volume 1, McGraw-Hill Book Company, 1953.

[17] T. Evans. The number of semigroup varieties, Quart. J. Math. Ser. (2) **19** (1968), 335-336.

[18] T. Evans. The lattice of semigroup varieties, Semigroup Forum **2** (1971), 1-43.

[19] C. F. Fennemore. All varieties of bands I, II, Math. Nachr. **48** (1971), 237-252.

[20] J. Fountain. Adequate semigroups, Proc. Edinburgh Math Soc. **22** (1979), 113-125.

[21] J. Fountain. Abundant semigroups, Proc. London Math. Soc. **44** (1982), 103-129.

[22] J. A. Gerhard. The lattice of equational classes of idempotent semigroups, J. Algebra **15** (1970), 195-224.

[23] M. Hall, Jr. "Theory of Groups", Chelsea Publishing Company, New York, 2nd ed., 1976.

[24] T. E. Hall. Representation, extension and amalgamation for semigroups, Quart. J. Math. Oxford Ser. (2) **29** (1978), 309-334.

[25] T. Hall, S. Kublanovsky, S. Margolis, M. Sapir, and P. Trotter. Decidable and undecidable problems related to finite 0-simple semigroups, J. Pure Appl. Algebra **119** (1997), 75-96.

[26] D. R. Hofstadter. "Gödel, Escher, Bach: An Eternal Golden Braid", Penguin Books Ltd., England, 1987.

[27] J. M. Howie. An embedding theorem with amalgamation for semigroups, *Proc. London Math. Soc.* (3), **12** (1962), 511-534.

[28] J.M.Howie. Semigroup amalgams whose cores are inverse semigroups, Quart. J. Math. Oxford Ser. (2) **26** (1975), 23-45.

[29] J.M. Howie. Fundamentals of Semigroup Theory, London Mathematical Society Monographs, Oxford University Press, 2nd ed., 1995.

[30] M. Jackson. Some undecidable embedding problems for finite semigroups, Proc. Edinburgh Math. Soc. **42** (1999), 113-125.

[31] M. Jackson. Finite semigroups whose varieties have uncountably many subvarieties, submitted for publication.

[32] M. Jackson. The embeddability of ring and semigroup amalgams is undecidable, submitted for publication.

[33] M. Jackson. On locally finite varieties with undecidable equational theory, submitted for publication.

[34] M. Jackson and O. Sapir. Finitely based sets of words, to appear in Internat. J. Algebra Comput..

[35] S. Jarman, N. Elliott, S. Nicol, A. McMinn and S. Newman. The base composition of the krill genome and its potential susceptibility to damage by UV-B, Antarctic Science **11** (1) (1999), 23-26.

[36] J. Jezek. Intervals in the lattice of varieties, Algebra Universalis **6** (1976), 147-158.

[37] J. Jezek. Nonfinitely based three-element idempotent groupoids, Algebra Universalis **20** (1985), 292-301.

[38] J. Kadourek. Uncountably many subvarieties of semigroups satisfying $x^2y \approx xy$, to appear in Semigroup Forum.

[39] O. G. Kharlampovich and M. V. Sapir. Algorithmic problems in varieties, Internat. J. Algebra Comput. **5** (1995), 379-602.

[40] E. I. Kleiman. On bases of identities of varieties of inverse semigroups, Sibirsk. Mat. Zh. **20** (1979), 760-777.

[41] D. J. Kleitman, B. R. Rothschild, and J.H. Spencer. The number of semigroups of order n, Proc. Amer. Math. Soc. **55** (1976), 227-232.

[42] I. O. Korjakov. A sketch of the lattice of commutative nilpotent semigroup varieties, Semigroup Forum **24** (1982), 285-317.

[43] V. Koubek and V. Rodl. Note on the number of monoids of order n, Comment. Math. Univ. Carolin. **26** (1985) 309-314.

[44] R. Kruse. Identities satisfied in a finite ring, J. Algebra **26** (1973), 298-318.

[45] S. Kublanovsky and M. V. Sapir. Potential divisibility in finite semigroups is undecidable, Internat. J. Algebra Comput. **8** (1998), 671-680.

[46] S. Kublanovsky and M. V. Sapir. A variety where the set of subalgebras of finite simple algebras is not recursive, Internat. J. Algebra Comput. **8** (1998), 681-688.

[47] I. V. L'vov. Varieties of associative rings I, Algebra i Logika **12** (1973), 269-297; II, 667-688.

[48] E. S. Lyapin. "Semigroups", Translations of mathematical monographs Volume 3, Amer. Math. Soc., Providence, RI, 1974.

[49] R. Lyndon. Identities in two valued calculi, Trans. Amer. Math. Soc. **71** (1951), 457-465.

[50] R. Lyndon. Identities in finite algebras, Proc. Amer. Math. Soc. **5** (1954), 8-9.

[51] R. McKenzie. A new product of algebras and a type reduction theorem, Algebra Universalis **18** (1984), 29-69.

[52] R. McKenzie. Tarski's finite basis problem is undecidable, Internat. J. Algebra Comput. **6** (1996), 49-104.

[53] G. McNulty and C. Shallon. Inherently nonfinitely based finite algebras, in; R. Freese and O. Garcia (eds.) Universal Algebra and Lattice Theory, Lecture Notes in Mathematics, vol 1149 (1983), Springer Verlag, Berlin, 205-231.

[54] A. Mekler, E. Nelson and S. Shelah. A variety with solvable but not uniformly solvable word problem, Proc. London Math. Soc. (3) **66** (1993), 225-256.

[55] M. Morse and G. Hedlund. Unending chess, symbolic dynamics, and a problem in semigroups, Duke Math J. **11** (1944), 1-7.

[56] V. L. Murskii. Concerning the number of k-element algebras with one binary operation which have no finite basis of identities, Problemy Kibernet **35** (1979), 5-27.

[57] V. L. Murskii. The existence of a finite basis and some other properties of "almost all" finite algebras, Problemy Kibernet **30** (1975), 43-56.

[58] H. Neumman. "Varieties of Groups", Springer Verlag, 1967.

[59] S. Oates and M. B. Powell. Identical relations in finite groups, J. Algebra **1** (1964), 11-39.

[60] J. Okniński and M. S. Putcha. Embedding finite semigroup amalgams, J. Austral. Math. Soc. Ser. A **51** (1991), 489-496.

[61] A. Ju. Ol'shanskii. On some infinite systems of identities, Amer. Math. Soc. Transl. Ser. 2 **119** (1983), 81-88.

[62] F. Pastijn. A representation of a semigroup by a semigroup of matrices over a group with zero, Semigroup Forum **10** (1975), 238-249.

[63] P. Perkins. Bases for equational theories of semigroups, J. Algebra **11** (1969), 298-314.

[64] P. Perkins. Basic questions for general algebra, Algebra Universalis **19** (1984), 16-23.

[65] M. Petrich. "Introduction to semigroups", Charles E. Merrill Publishing Co. 1973.

[66] D. Pigozzi. On some operations on classes of algebras, Algebra Universalis **2** (1972), 346-353.

[67] J. E. Pin. "Varieties of Formal Languages", North Oxford Academic Publishers Ltd (English edition), 1986.

[68] G. Pollak. On hereditarily finitely based varieties of semigroups, Acta Sci. Math. **37** (1975), 339-348.

[69] G. Pollak. On identities that define hereditarily finitely based varieties of semigroups, in "Algebraic Theory of Semigroups", Proc. Conf. Szeged (1976), Colloq. Math. Soc. János Bolyai, North-Holland Amsterdam **20** (1979), 447-452.

[70] G. Pollak and M. V. Volkov. On almost simple semigroup identities, in "Semigroups", Proc. Conf. Szeged (1981), Colloq. Math. Soc. János Bolyai, North-Holland Amsterdam **39** (1985), 287-323.

[71] V. V. Rasin. Varieties of orthodox Clifford semigroups, Russian Math. (Izv. VUZ) **26** No. 7 (1982), 82-85.

[72] J. J. Rotman. "An Introduction to the Theory of Groups," Springer Verlag, 4th ed., 1991.

[73] M. V. Sapir. Problems of Burnside type and the finite basis property in varieties of semigroups, Math. USSR Izv. **30** No. 2 (1988), 295-314.

[74] M. V. Sapir. Inherently nonfinitely based finite semigroups, Russian Acad. Sci. Sb. Math. **61** No.1 (1988), 155-166.

[75] M. V. Sapir. On Cross semigroup varieties and related questions, Semigroup Forum **42** (1991), 345-364.

[76] M. V. Sapir. Eventually $\mathcal{H}$-related sets and systems of equations over finite semigroups and rings, J. Algebra **183** (1996), 365-377.

[77] M. V. Sapir. Algorithmic problems for amalgams of finite semigroups, manuscript (1999).

[78] M. V. Sapir and M. V. Volkov. HFB property and structure of semigroups, Contri- butions to General Algebra, Vol.6, Vienna (1988), 303-310.

[79] O. Sapir. Identities of finite semigroups and related questions, PhD thesis, University of Nebraska, 1997.

[80] O. Sapir. Finitely Based Words, to appear in Internat. J. Algebra Comput..

[81] O. Schreier. Die untergruppen der freien gruppen, Abhandlungen aus dem mathematischen Seminar der hansischen Universität Hamburg **5** (1927), 161-183.

[82] L. N. Shevrin and M. V. Volkov. Identities of semigroups, Izv. Vyssh. Uchebn. Zaved. Mat. (1985), No. 11, 3-47.

[83] L. N. Shevrin. Quasiperiodic semigroups decomposable into a band of Archimedean semigroups, Sixteenth All-Union Algebra Conf., Abstracts of Reports, Part 1, Leningrad. Otdel. Mat. Inst. Akad. Nauk. SSSR, Leningrad (1981), 177-178.

[84] L. N. Shevrin. To the theory of epigroups I, Matem. Sbornik **185** No. 9 (1994), 129-160.

[85] L. N. Shevrin. To the theory of epigroups I, Matem. Sbornik **185** No. 8 (1994), 153-176.

[86] L. M. Shneerson. On the axiomatic rank of varieties generated by a semigroup or monoid with one defining relation, Semigroup Forum **39** (1989), 17-38.

[87] P. J. Smith. "Into Statistics", Thomas Nelson Australia, 1993.

[88] R. R. Stoll. "Set theory and logic", W. H. Freeman and Company, 1963.

[89] H. Straubing. The variety generated by all finite nilpotent monoids, Semigroup Forum **24** (1982), 25-38.

[90] A. V. Tishchenko. The finiteness of a base of identities for five element monoids, Semigroup Forum **20** (1980), 171-186.

[91] A. N. Trahtman. A variety of semigroups without an irreducible basis of identities, Mat. Zametki **21** (1977), 865-872.

[92] A. N. Trahtman. Six-element semigroup generates a variety with uncountably many subvarieties, Alg. Systems and their varieties, Sverdlovsk (1988), 138-143 (Russian).

[93] A. N. Trahtman. Identities of a five-element 0-simple semigroup, Semigroup Forum **48** (1994), 385-387.

[94] M. R. Vaughan-Lee. Uncountably many varieties of groups, Bull. London Math. Soc. **2** (1970), 280-286.

[95] B. Wells. "Pseudorecursive varieties and their implications for word problems", Doctoral Dissertation, University of California, Berkeley, 1983.

[96] B. Wells. Pseudorecursive varieties of semigroups - I, Internat. J. Algebra Comput. **6** (1996), 457-510.

[97] A. I. Zimin. Blocking sets of words, Mat. Sb. **119** (1982), 363-375.

Index