Theory and Applications of Categories, Vol. 18, No. 11, 2007, pp. 303–320.

THE THEORY OF CORE ALGEBRAS: ITS COMPLETENESS

PETER FREYD

Abstract. The core of a category (first defined in "Core algebra revisited" *Theoretical Computer Science,* Vol 375, Issues 1–3 pp 193–200) has the structure of an abstract core algebra (first defined in the same place). A question was left open: is there more structure yet to be defined? The answer is no: it is shown that any operation on an object arising from the fact that the object is the core of its category can be defined using only the constant and two binary operations that appear in the definition of abstract core algebra. In the process a number of facts about abstract core algebras must be developed.

Let $\mathbb{A}$ be a category with finite products. A CORE of $\mathbb{A}$—if such exists—is an object C together with a transformation $(-) \times C \longrightarrow (-)$ which transformation is universal among such: that is, any other natural transformation of the form $(-) \times A \longrightarrow (-)$ is induced by a unique map $A \longrightarrow C$.

C comes equipped with a monoid structure, that is, it has a "constant" $1 \longrightarrow C$ and a binary operation $C \times C \longrightarrow C$ satisfying the axioms for a monoid. $e : 1 \longrightarrow C$ is defined as the unique map such that

$$\begin{tikzcd}
(-) \times 1 \arrow[r, "1 \times e"] \arrow[dr, "\cong"'] & (-) \times C \arrow[d, "t"] \\
& (-)
\end{tikzcd}$$

where t is the defining transformation for the core. The multiplication $m : C \times C \longrightarrow C$ is defined as the unique map such that

$$\begin{tikzcd}
(-) \times (C \times C) \arrow[rr, "1 \times m"] \arrow[d, "a"'] & & (-) \times C \arrow[d, "t"] \\
((-) \times C) \times C \arrow[r, "t \times 1"] & (-) \times C \arrow[r, "t"] & (-)
\end{tikzcd}$$

where a is the associativity isomorphism.

Received by the editors 2007-01-26 and, in revised form, 2007-07-14.

Transmitted by Michael Barr. Published on 2007-07-22.

2000 Mathematics Subject Classification: 18A40.

Key words and phrases: core, cored category, abstract core algebra, critical lemma, no-lost-variables, base cancellation .

If we notate things as if there were elements and denote the values of the canonical transformation at X by $t_X\langle x, c\rangle = x\uparrow c$, then we have defined 1 so that $x\uparrow 1 = x$ and we have defined the product so that $x\uparrow(cd) = (x\uparrow c)\uparrow d$.

Besides the monoid structure on C there is another binary operation on C, to wit, $t_C\langle x, y\rangle = x\uparrow y$. We have an object with a constant and two binary operations satisfying the equations of an (abstract) CORE ALGEBRA:

1, 1′:	$1x = x = x1,$
2:	$x(yz) = (xy)z,$
3, 3′:	$1\uparrow x = 1,\ x\uparrow 1 = x,$
4:	$(xy)\uparrow z = (x\uparrow z)(y\uparrow z),$
5:	$x\uparrow(yz) = (x\uparrow y)\uparrow z,$
6:	$xy = y(x\uparrow y),$
7:	$(x\uparrow y)\uparrow z = (x\uparrow z)\uparrow(y\uparrow z).$

An alternate notation (but one that clashes with standard notation when discussing M-sets) is to denote $x\uparrow y$ as x^y. The equations rewrite as:

1, 1′:	$1x = x = x1,$
2:	$x(yz) = (xy)z,$
3, 3′:	$1^x = 1,\ x^1 = x,$
4:	$(xy)^z = x^z y^z,$
5:	$x^{yz} = (x^y)^z,$
6:	$xy = yx^y.$

Equation 7 is best rewritten as

$$(x\uparrow y)^z = x^z\uparrow y^z$$

Equations $1', 3'$ and 7 are present for aesthetic reasons; they are, in fact, redundant. In my paper "Core algebra revisited" [1] it was shown that any equation that holds for all categorical cores is a consequence of equations 1 through 6. (Indeed, any universal first order sentence that holds for all categorical cores holds for all abstract cores.)

What was not proven in "Core algebra revisited" is the completeness of the structure itself:

1. THEOREM. *The theory of abstract core algebras is the complete algebraic structure enjoyed by categorical cores.*

Put another way: any operator that can be defined on an object using only the fact that it is the core of a category is already definable from 1, xy and $x\uparrow y$.

The proof we have found requires far more work than expected.

Matters would be simplified if we could find the "universal cored category", that is, the initial object in the category of cored categories. (By the later we mean the category whose object are categories—small, if you insist—with finite products and a core and

[1] *Theoretical Computer Science* (Vol 375, Issues 1–3 pp 193–200)—available online to subscribers

whose maps are natural equivalence classes of functors that preserve finite products and cores.) Then the algebraic structure of the core in the universal cored category is the answer to the question. We would need to show that it is precisely the theory of abstract core algebras as just defined.

A proof looks straightforward. Let $\mathbb{C}$ be the "Lawvere category" of the theory of core algebras (Lawvere would call it the theory). Its objects are the finite cartesian powers, $1,\ C,\ C\times C, \ldots,\ C^n, \ldots,$ of an abstract object C. The maps from C^m to C^n are known when we know the maps from C^m to C (because C^n is a cartesian power) and the maps C^m to C are named by terms built from the constant and two binary operations of the theory of core algebras, two terms naming the same map iff the defining equations force them to.[2]

Quite clearly $\mathbb{C}$ has a functor, unique up to natural equivalence, to every cored category, $\mathbb{A}$, which functor preserves finite products and carries the abstract object C to a core of $\mathbb{A}$.

Alas, $\mathbb{C}$ is not a cored category. There is, indeed, a distinguished natural transformation $(-)\times C \longrightarrow (-)$. Because the target of the transformation (the identity functor) preserves products a transformation $C^n \times A \longrightarrow C^n$ is known once it is known when $n = 1$. Equations 3, 4 and 7 are just what are needed to know that the map $C\times C \longrightarrow C$ denoted by $x\,{\uparrow}\,y$ induces a natural transformation—in elemental notation, for any n-term ϕ it is the case that $(\phi\langle x_1, x_2, \ldots, x_n\rangle)^y = \phi\langle x_1^y, x_2^y, \ldots, x_n^y\rangle$.

But this distinguished transformation is not universal in $\mathbb{C}$. The theory of core algebras has a unique constant, 1. There are no further derived constants (in traditional language, equations 1 and 3 say that the one-element set $\{1\}$ is a sub-algebra of any core algebra). Hence there is a unique map $1 \longrightarrow C$, consequently a unique map from 1 to any object in $\mathbb{C}$. That is, the terminator in $\mathbb{C}$ is a co-terminator; $\mathbb{C}$ is a "punctuated category", a "category with zero". But

2. Proposition. *If the terminator of a cored category is a co-terminator then the category is equivalent to the one-object one-morphism category.*

Because: the maps of the form

$$X \times 1 \longrightarrow 1 \longrightarrow X$$

comprise a natural transformation, necessarily induced by a map $1 \longrightarrow C$; but there's only one map $1 \longrightarrow C$ and it induces the projection $(-)\times 1 \longrightarrow (-)$; the only way these natural transformations can be equal is if all objects are isomorphic to 1.

Since $\mathbb{C}$ is not degenerate (that is, the theory is consistent: one can not prove $x = y$) it can not be a cored category. (In fact the category of cored categories does not have an initial object.)

[2] To be more formal: a map from C^m to C is named by a sequence of m variables and a term thereon; another such sequence and term name the same map if after an appropriate substitution of variables (so that the two sequences are the same) the resulting terms are provably equal. Composition of maps is effected by substitution (with care being taken not to confuse variables).

The fact that $\mathbb{C}$ is a punctuated category obstructs the proof. One way to remove the obstruction is to adjoin a "generic object" to $\mathbb{C}$. To that end, consider the two-sorted theory where the two sorts are a core algebra and an object on which it acts. (This process is similar to the passage from the one-sorted theory of a monoid to the two-sorted theory of a monoid together with a set on which it acts.) The Lawvere category, $\mathbb{C}[X]$, has a pair of base objects X and C; all the objects are cartesian products of these: $X^m \times C^n$. C has the structure of a core algebra and X comes equipped with a special map $X \times C \longrightarrow X$. As is the case with monoids and sets on which they act (and with rings and modules) we will "overload" the elemental notation: if we understand $x \in X$ and $c \in C$ then $x \uparrow c \in X$ describes the action $X \times C \longrightarrow X$. Besides the equations for core algebras we need the two equations $x \uparrow 1 = x$ and $(x \uparrow c) \uparrow d = x \uparrow (cd)$. The analog of equation 7 is a consequence: $(x \uparrow c) \uparrow d = x \uparrow (cd) = x \uparrow (d(c \uparrow d)) = (x \uparrow d) \uparrow (c \uparrow d)$; that is, for any term ϕ in this new two-sorted theory we still have $(\phi\langle x_1, x_2, \ldots, x_n\rangle)^y = \phi\langle x_1^y, x_2^y, \ldots, x_n^y\rangle$.

Alas, $\mathbb{C}[X]$ is not a cored category. Note that there is only one map from X to X (the identity map) and only one from X to C (the constant map); that is, X is a co-terminator. But

3. Proposition. *If a cored category has a co-terminator,* 0, *then it is a* strict *co-terminator, that is, any map targeted at* 0 *is an isomorphism.*[3]

Because: since there can be only one transformation from $(-) \times 0$ to $(-)$ (there's only one map $0 \longrightarrow C$) we conclude that the left-projection $(-) \times 0 \longrightarrow (-)$ is the same as the right-projection followed by the unique map from the co-terminator $(-) \times 0 \longrightarrow 0 \longrightarrow (-)$; thus if there's a map from Y to 0 there's a factorization of 1_Y as $Y \longrightarrow Y \times 0 \longrightarrow 0 \longrightarrow Y$; as always for co-terminators $0 \longrightarrow Y \longrightarrow Y \times 0 \longrightarrow 0$ is a factorization of 1_0, all of which makes $Y \longrightarrow 0$ an isomorphism. We could, of course, re-define the category to make X into a strict co-terminator, but the result would be the same as adjoining a strict co-terminator to the first attempt and any category with two endomorphisms of its identity functor still has those endomorphisms after adjoining a strict co-terminator; the previous argument would still apply.

The next attempt is to adjoin two generic objects, $\mathbb{C}[X, Y]$. Alas the pathology can be reconstructed, indeed any finite number of generic objects will fail: the finite product, P, of those objects would be a co-terminator. The structure might indeed be a cored category but it would not have a product- and core-preserving functor to every other cored category. Functors that preserve finite products and cores need not preserve co-terminators, but they would have to preserve the fact that the left and right projections from $P \times P$ to P are equal, that is, they would have to preserve the fact that P is a sub-terminator. For any group, G, the category of G-sets is a cored-category (as was shown in "Core algebra revisited") and easily remains such if the empty G-set is removed. Since the only sub-terminators left have only one element P would have to be sent to a one-element G-set. Worse, so would $P \times C$. When G is non-trivial C is non-trivial (it's the "conjugacy" G-set) and the functor could not be core-preserving.

[3]The previous proposition is, of course, an immediate corollary of this one.

But not all is lost:

4. LEMMA. *The category* $\mathbb{C}[X_1, X_2, \ldots, X_n, \ldots]$ *obtained by adjoining an infinite sequence of generic objects on which* $\mathbb{C}$ *acts is a cored category and is weakly initial in the category of cored categories.*

This fact implies the main result since the complete algebraic structure of the core in $\mathbb{C}[X_1, X_2, \ldots, X_n, \ldots]$ is just the theory of abstract core algebras and it has a product- and core-preserving functor into any cored category.[4] Given a recipe for, say, an n-ary operation on cores, and assuming that the recipe is preserved by product- and core-preserving functors, then in any particular cored-category $\mathbb{A}$ we can choose a product- and core-preserving functor, $T : \mathbb{C}[X_1, X_2, \ldots, X_n, \ldots] \longrightarrow \mathbb{A}$. The recipe delivers a map in $\mathbb{C}[X_1, X_2, \ldots, X_n, \ldots]$ that can be obtained from the abstract core-algebra structure on the core. Hence, so must be the case in $\mathbb{A}$.

As already mentioned there is a lot of work to do. There are two lemmas that will, in particular, require work. One is the "No Lost Variables" lemma in the statement of which the phrase REDUCED CORE-ALGEBRA TERM means only that the constant 1 does not appear.

5. LEMMA. [The No-Lost-Variables Lemma] *When a reduced core-algebra term is interpreted as a reduced word in a free group (with its unique core-algebra structure), each variable that appears in the reduced core-algebra term continues to appear in the reduced word.*

The unique core-algebra structure on a group takes $x \uparrow y$ as $y^{-1}xy$ (as forced by defining equation 6).

This lemma is the most difficult ingredient in the proof of this partial converse to equation 6:

6. LEMMA. [The Critical Lemma] *If* ϕ *and* ψ *are core-algebra terms,* v *a variable which does not appear in* ϕ *and if* $v\phi = \phi\psi$ *then* $\psi = v \uparrow \phi$.

(Note the importance of the condition: if omitted take $\phi = \psi = v$ to obtain $v\phi = \phi\psi$ with $\psi \neq v \uparrow \phi$.)[5]

We'll also need

7. LEMMA. [The Base-Cancellation Lemma] *If* v *is a variable which does not appear in the terms* ϕ *and* ϕ' *and if* $v \uparrow \phi = v \uparrow \phi'$ *then* $\phi = \phi'$.

We must show that for any object $A = X_{a_1} \times X_{a_2} \times \cdots X_{a_k} \times C^m$ in $\mathbb{C}[X_1, X_2, \ldots, X_n, \ldots]$ and transformation $(-) \times A \longrightarrow (-)$ there is a unique map $A \longrightarrow C$ that induces it.

Finding the map is easy. Let n be different from any of the a's. Then $X_n \times A \longrightarrow X_n$ is named by an X_n-valued term whose input variables are from the objects involved in

[4] The weakness, note, is extreme: unless the target category is equivalent to the degenerate category there will be $2^{\aleph_0}$ non-equivalent product- and core-preserving functors.

[5] A fairly easy induction shows that for any term, θ, on one variable, v, it is the case that $vv^n = v^n(v \uparrow \theta)$ for all sufficiently large powers v^n.

$X_n \times A$. It is easily to check that the only such terms are necessarily of the form $x_n \upharpoonright \phi$ where ϕ is a C-valued term. And it is easy to check that a C-valued term can not depend on any variables other than the C-terms. Put another way, the map $A \longrightarrow C$ named by ϕ is of the form $A \longrightarrow C^m \longrightarrow C$ where $A \longrightarrow C^m$ is the projection.

We must show that $C \times A \longrightarrow C$ is named by the term $v \upharpoonright \phi$ and for any j that $X_j \times A \longrightarrow X_j$ is named by the term $x_j \upharpoonright \phi$ (where x_j names the left projection $X_j \times A \longrightarrow X_j$).

To that end let ψ be a term that names $C \times A \longrightarrow C$. We wish to show $\psi = v \upharpoonright \phi$. As above, ψ can not depend on any variables other than those in C. The naturality of the transformation implies the commutativity of:

$$\begin{tikzcd}
(X_n \times C) \times A \arrow[r] \arrow[d, "a \times 1"'] & X_n \times C \arrow[d, "a"] \\
X_n \times A \arrow[r] & X_n
\end{tikzcd}$$

where a denotes the action of C on X_n, that is, the map named by the term $x_n \upharpoonright v$.

Traveling counter-clockwise we obtain the map named by the term $(x_n \upharpoonright v) \upharpoonright \phi$ and traveling clockwise, $(x_n \upharpoonright \phi) \upharpoonright \psi$. That is, $x_n \upharpoonright (v\phi) = x_n \upharpoonright (\phi\psi)$. The equality of these terms implies via the Base-Cancellation Lemma (specialize the generic object to C) that $v\phi = \phi\psi$. The critical lemma then yields $\psi = v \upharpoonright \phi$.

For j such that X_j is not involved in A we let ϕ' denote the term such that $x_j \upharpoonright \phi'$ describes the natural transformation $(-) \times A \longrightarrow (-)$ when $(-) = X_j$. We wish to show $\phi' = \phi$. The same diagram as above with n replaced by j yields the equation $x_j \upharpoonright (v\phi') = x_j \upharpoonright (\phi'\psi)$ hence $\psi = v \upharpoonright \phi'$. Base-cancellation then yields $\phi = \phi'$.

Finally, for j such that X_j is involved in A we must dispel the possibility that the term that describes the transformation $(-) \times A \longrightarrow (-)$ when $(-) = X_j$ depends on some X_j-variable from A. We'll let x_j continue to denote the variable that names the left-projection $X_j \times A \longrightarrow X_j$ and y_j will name a map of the form $X_j \times A \longrightarrow A \longrightarrow X_j$ where $X_j \times A \longrightarrow A$ is the right-projection. What we must dispel is the possibility that the transformation is described by a term of the form $y_j \upharpoonright \phi'$. Using the same diagram we have that $y_j \upharpoonright \phi'$ describes the counter-clockwise map and $(y_j \upharpoonright \phi') \upharpoonright \psi$ describes the clockwise map. But we have that $\psi = v \upharpoonright \phi$ hence the equation $\phi' = \phi'(v \upharpoonright \phi)$ where v does not appear as a variable in either ϕ or ϕ'. By instantiating all the other variables as 1 we obtain $1 = v$, a contradiction.

We have thus reduced the completeness of the theory of core algebras to the Critical Lemma. No categorical considerations remain; the rest is syntax.

Proof of the Critical Lemma. We wish to prove:

If ϕ and ψ are core-algebra terms, if v is a variable which does not appear in ϕ and if $v\phi = \phi\psi$ then $\psi = v \upharpoonright \phi$.

We will show: first, that ψ is of the form $v \upharpoonright \theta$; second that v does not appear in θ; and, finally, using that θ is v-free, that $\phi = \theta$.

In the course of the proof we will find it useful to map the free core algebra into various special core algebras, which maps will be called "specializations". There will be four of them. The first and third will land in familiar places: the first will be the free commutative monoid with its unique core-algebra structure ($x \uparrow y = x$); the third will be the free group with its unique core-algebra structure ($x \uparrow y = y^{-1}xy$). The second and fourth specializations are exotic. Their construction will be much aided by introducing the notion of a SEMI-CORE ALGEBRA, the result of removing the constant 1 from the definition and retaining just the four equations $(2, 4, 5, 6)$ in which 1 does not appear. Given a semi-core algebra we may adjoin a new element called 1 and extend the definition of the two binary operations in the unique way that satisfies equations 1, $1'$, 3, $3'$. The result is always a core algebra: a straightforward (but tedious) inspection shows that whenever any one of the variables is instantiated with 1, the semi-core equations automatically hold.[6]

The first specialization will be used to show that ψ is of the form $v \uparrow \theta$. The second and third specializations will be used to show that v does not appear in θ. The fourth specialization will capitalize on the v-freeness of θ to establish $\phi = \theta$.

It would be much simpler if we had a "canonical form" theorem for core-algebra terms. There's no such theorem in sight.

There's no canonical form but we can define a NORMAL CORE-ALGEBRA TERM recursively by stipulating that 1) every variable v is a normal term, 2) if ϕ is normal then $v \uparrow \phi$ is normal, and 3) a product of any finite sequence of reduced normal terms is normal. (We will understand that 1 is the product of the empty sequence.)

It isn't hard to see that every term is provably equal to a normal term, but we want a bit more, to wit, no lost variables. So define a "normalization" process on reduced terms inductively by taking

$$N(v) = v;$$

$$N(\phi_1\phi_2) = N(\phi_1)N(\phi_2);$$

$$N(\phi_1 \uparrow \phi_2) = v_1^{\theta_1 N(\phi_2)} v_2^{\theta_2 N(\phi_2)} \cdots v_n^{\theta_n N(\phi_2)}$$

where the θ's are normal and $N(\phi_1) = v_1^{\theta_1} v_2^{\theta_2} \cdots v_n^{\theta_n}$. An easy induction shows that any variable in a reduced term ϕ appears also in $N(\phi)$.

We'll call the v's in a normal term such as $v_1^{\theta_1} v_2^{\theta_2} \cdots v_n^{\theta_n}$ the BASE VARIABLES.

8. LEMMA. *The "multiset" (set-with-multiplicities) of base variables of a normal term is invariant.*

Because: when we specialize to the free commutative monoid with its unique core-algebra structure; a normal term is sent to the product of its base variables. (Note that the order in which the base variables appear is not at all unique, indeed, one may use the defining equation 6 to reorder them any way one chooses.)

---

[6] Indeed, this step uses only $1, 1', 3, 3'$, that is, whenever any variable is replaced with a 1 those four equations imply the other four, 2,4,5,6. (We used this fact in the independence examples in "Core algebra revisited.")

We thus simplify the critical lemma. Suppose $v\phi = \phi\psi$ and v does not appear in ϕ. In this first of the four specializations, we apply the reflector from the free core algebra to the free commutative monoid to see that the normalization of ψ has a unique base variable, namely v. When ψ is normalized it is of the form $v \uparrow \theta$ where θ is normal.[7] So it suffices to prove:

If ϕ and θ are core-algebra terms, v a variable which does not appear in ϕ and if $v\phi = \phi(v \uparrow \theta)$ then $\theta = \phi$.

The next step is to show that v can not appear in θ.

The second specialization will be used to dispatch the possibility that v appears as a base variable in θ. We will specialize to a core algebra with just three elements $1, v$ and 0. It is constructed most easily by starting with the two-element semi-core algebra $\{v, 0\}$ in which the value of both binary operations is taken to be constantly 0. (The result is a semi-core algebra because none of the defining equations for semi-core algebras is singular, that is, because each side of each of the four equations has a binary operation, hence will automatically be valued as 0.) The resulting 3-element core algebra [8] is given by:

$$\begin{array}{c|ccc} xy & 1 & v & 0 \\ \hline 1 & 1 & v & 0 \\ v & v & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \qquad\qquad \begin{array}{c|ccc} x \uparrow y & 1 & v & 0 \\ \hline 1 & 1 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}$$

The specialization we want carries v to v and all other variables to 1. It carries $v\phi$ to v. If v is a base variable in θ then θ will be carried to either v or 0. In either case, $v \uparrow \theta$ will be carried to 0 and so will $\phi(v \uparrow \theta)$. That is, if v appears as a base variable in θ then the equation $v\phi = \phi(v \uparrow \theta)$ becomes $v = 0$, a contradiction.

The third specialization is to the free group. We will show that if $v\phi = \phi(v \uparrow \theta)$ where v does not appear in ϕ and is not a base variable in θ then ϕ and θ become equal when specialized to the free group. In particular, then, v does not appear in the specialization of θ. Coupled with the No-Lost-Variables Lemma (proven below) such implies that v does not appear in θ before specialization.

Let w denote the reduced word obtained when θ is specialized to the free group. Because v is not a base variable in θ we know that the "total degree" of v in w is 0 (a formal definition of the total degree of v is the power of v that is obtained when all other variables are replaced by 1). Let u denote the reduced word to which ϕ is specialized. Then the equation $v\phi = \phi(v \uparrow \theta)$ becomes the equation $vu = uw^{-1}vw$ which rewrites to $v(uw^{-1}) = (uw^{-1})v$. That is, uw^{-1} commutes with v in the free group, hence must be a power of v. The hypotheses on ϕ and θ tell us that the power in question is $v^0 = 1$. Hence $u = w$.

The No-Lost-Variables Lemma thus reduces the Critical Lemma to:

---

[7] An extra consequence is that when v is replaced with 1 then ψ reduces to 1. Thus if $v\phi = \phi'\psi$ and v does not appear in either ϕ or ϕ' then necessarily $\phi = \phi'$.

[8] It is one of the 16 isomorphism types of 3-element core algebras. There are 3 types of 2-element core algebras. See addendum at end.

If ϕ and θ are core-algebra terms, v a variable which appears in neither ϕ nor θ and if $v\phi = \phi(v\uparrow\theta)$ then $\theta = \phi$.

For the fourth specialization we use the following construction. Let S be a semi-core algebra, and v an element not in S. Create a new semi-core algebra, S', by enlarging S to include the element v and a disjoint copy of S whose elements will be written in the form vx where $x \in S$. The semi-core algebra structure on S' is given by:

$\star$	y	v	vy
x	$x \star y$	v	vy
v	vy	v	vy
vx	$v(xy)$	v	vy

where $\star$ denotes either of the two binary operations. (Note that $(vx)\uparrow y = v(xy)$.)

The verification of the semi-core equations is a bit tedious. Clearly S appears as a sub-algebra, hence we needn't bother with the case that v doesn't appear in any of the variables. Note that in each of the four equations the right-most variable is the same on the two sides and that easily dispatches the case where the right-most variable is of the form v or vz. We may henceforth assume that the right-most variable is v-free. Equation 6 then becomes $(vx)y = y((vx)\uparrow y)$ and it is right here that we use $(vx)\uparrow y = v(xy)$. The only other equation that gives any pause is 4 in the case that the first variable is of the form vx and the other two are v-free: $((vx)y)\uparrow z = vxyz$ and $((vx)\uparrow z)(y\uparrow z) = (vx)z(y\uparrow z)$. Equation 4 thus holds for S' because Equation 6 holds for S.

The Critical Lemma is now easily dispatched. Using that v appears in neither ϕ nor θ, we take S to be a free semi-core algebra that contains both ϕ and θ but not v. The specialization to the resulting core algebra sends $v\phi$ to $v\phi$ and $\phi(v\uparrow\theta)$ to $v\theta$. If ϕ and θ were distinct, then so would be $v\phi$ and $v\theta$. [9]

The Base-Cancellation Lemma is quickly dispatched using the same specialization.[10]

Proof of The No-Lost-Variables Lemma. Recall that a reduced core-algebra term is one in which the constant 1 does not appear (in particular all normal terms other than 1 are reduced).

When a reduced core-algebra term is interpreted as a reduced word in a free group each variable that appears in the reduced core-algebra term continues to appear in the reduced word.

Let $\phi = v_1^{\theta_1} v_2^{\theta_2} \cdots v_n^{\theta_n}$. be a normal core-algebra term and suppose we know the lemma is correct for all smaller normal terms, in particular for the θ's. Let $w_1, w_2, \ldots, w_n$ be the reduced words that result when the θ's are interpreted in a free group. From here on it is

[9] Note that, in fact, we proved a stronger lemma: *If ϕ, ϕ' and θ are core-algebra terms, v a variable which appears in neither ϕ, ϕ' nor θ and if $v\phi = \phi'(v\uparrow\theta)$ then $\theta = \phi$.*

[10] This is a deeply unsatisfactory proof: the condition of v-freeness appears quite unnecessary. In an appendix we sketch a possible (but quite unpleasant) proof: it argues that any proof of $v\uparrow\phi = v\uparrow\phi'$ can be converted into a proof of $\phi = \phi'$ whether or not v appears in ϕ or ϕ'.

an exercise in the syntax of groups—no more core algebras. We need to show that every variable occurring in $v_1^{w_1} v_2^{w_2} \cdots v_n^{w_n}$ continues to occur even after its reduction.

We start with the unreduced catenation, u, of the reduced words

$$w_1^{-1}, v_1, w_1, w_2^{-1}, v_2, w_2, \ldots, w_n^{-1}, v_n, w_n$$

We'll refer to the letters in u that correspond to the v's as "base letters".

We construct a planar diagram by first placing a horizontal row of evenly placed "nodes", $|u|$ of them, where $|u|$ is the length of u. Each node comes equipped with an assigned letter and an assigned signature. We'll call the nodes that correspond to the base letters the "base nodes".

Next, for each i and each letter in w_i add a semicircle below the row of nodes, the two ends of which are the two nodes corresponding to the two positions of the letter, one in w_i^{-1}, the other (to the right) in w_i. Thus for each i there will be a family of $|w_i|$ concentric semicircles, where $|w_i|$ is the length of w_i. Their common center is the ith base node.

We're about to add semicircles above the row of nodes. The ones we've just added below, we'll call the "lower" semicircles.

Add the "upper" semicircles by iterating the following (non-deterministic) procedure. Attach a new upper semicircle to a pair of nodes if the following four conditions are met: 1) each node in the pair is not yet attached to an upper semicircle; 2) they have the same assigned letter; 3) they have opposite signatures; 4) all the nodes between the pair are themselves paired off as ends of previously attached upper semicircles. (Hence the first such pair of nodes must be adjacent.) Continue as long as possible, that is, until no such pair of nodes can be found.

We'll call a node that remains unattached to an upper semicircle "unreduced". The sequence of unreduced nodes corresponds to a sequence of letters and signatures that describes the reduced word of interest. We need to show that each letter appearing anywhere in u appears as the letter assigned to an unreduced node.

A few easy observations: the letters assigned to the ends of each semicircle, upper or lower, are the same; the signatures are opposite; below any upper circle there are no unreduced nodes; any node is attached to at most two semicircles, one upper, one lower; the only nodes not attached to a lower semicircle are base nodes.[11]

The diagram falls apart as a disjoint union of connected components. The components come in three varieties: 0) closed paths 1); single nodes; and 2) open paths with two ends. The semicircles in the open and closed paths alternate between upper and lower and as do the the signatures of the nodes.

At least one end of any open path is an unreduced node.

Suppose not. Since each end is attached to an upper semicircle, neither is attached to a lower semicircle and hence, each must be a base node. The signature of any base node is

[11] Another easy observation is that each lower semicircle has an odd number of nodes between its ends and an upper semicircle an even number. We know no use for this observation. (At least it tells us that there are no circles.)

positive. Because the signatures alternate, the component necessarily has an odd number of nodes, hence an even number of semicircles. We obtain a contradiction: since the semicircles also alternate, this time between upper and lower, the evenness of the number of semicircles forces one of the end semicircles to be lower; that end is necessarily an unreduced node.[12]

There are no closed paths.

Because, given a closed path, note first that there can be no unreduced nodes inside the path (if there were, attach a vertical line through it; the line must meet both an upper and lower semicircle of the closed path; the upper semicircle would then have an unreduced node below it). Hence inside any closed path there can be no isolated nodes or open paths; within any closed path there are only closed paths.

Hence, if there were any closed path there would have to be an innermost closed path, that is, a path with no nodes inside of it. Consider the rightmost node, x, on an innermost path. Let y be the node just to its left. Necessarily y is also on the closed path. Since x is rightmost, the lower semicircle attached to it arcs to the left, hence the lower semicircle attached to y is concentric. That is x and y correspond to adjacent letters on w_i for some i. Add an arrow to each semicircle so that when one crosses the semicircle traveling from inside to outside, the arrow is pointed to the right. The lower semicircle attached to x thus points away from x and the lower semicircle attached to y points towards y. Thus if we travel from x to y following the arrows we will reach y via the lower semicircle attached to it. This open path connecting x and y thus has an odd number of semicircles and the number of its nodes is even. The signatures attached to x and y must be opposite. The same letter is assigned to each (they're connected by a path). All of which contradicts the fact that w_i is a reduced word. That is, an innermost closed path can not have a rightmost node. Hence there is no innermost closed path. Hence no closed paths at all.

Hence,

If there's a node assigned to a letter then there's an unreduced node assigned to the same letter.

Because given a node consider the component that contains it; all nodes on that component are assigned to the same letter and, since it is not a closed path, at least one of its two ends is unreduced.

Appendix

9. Proposition. [Unrestricted Base-Cancellation] *If v is a variable, ϕ and ϕ' terms (with or without v) and if $v \uparrow \phi = v \uparrow \phi'$ then $\phi = \phi'$.*

[12] In fact, each base variable is either an isolated node or appears as an end of an open path whose other end is unreduced and has positive signature; all other open paths have unreduced nodes at each end and they have opposite signatures.

For technical reasons we need the notion of a PSEUDO-CORE ALGEBRA (that is, for reasons other than the existence of interesting examples): it is obtained by removing equation 6 from the axioms and restoring $1'$ and $3'$:

1, $1'$:	$1x = x = x1$,
2:	$x(yz) = (xy)z$,
3, $3'$:	$1 \uparrow x = 1,\ x \uparrow 1 = x$,
4:	$(xy) \uparrow z = (x \uparrow z)(y \uparrow z)$,
5:	$x \uparrow (yz) = (x \uparrow y) \uparrow z$.

The normal forms described above now do turn out to be canonical forms. That is, the free pseudo-core algebra on a set of letters is precisely the set of normal terms that can be written with those letters.

We find it useful to view these terms as trees, to wit, the set of rooted planar trees with branch-labels. The word "planar" means that the branches from each node come with a total ordering (which we will read from left to right). The root will be where nature intended: at the bottom. The branch-labels are restricted to the given set of letters.

For a more formal approach, define a tree to be a finite sequence of the form

$$a[A]b[B]\cdots z[Z]$$

where $a, b, \ldots, z$ are letters and $A, B, \ldots, Z$ are themselves trees (hence a tree is a sequence of letters and the symbols "[","]" subject to certain rules).[13]

We impose the structure of a pseudo-core algebra on the set of trees by taking the branchless rooted tree to be 1; by defining the product of a pair of trees ST to be the result of joining them at the root; and by defining $S \uparrow T$ to be the result of joining the root of a copy of T to each node of S that appears at the top of a branch attached to the root. The defining equations are easily verified.

In the more formal approach 1 is the empty sequence, product is catenation, and

$$(a[A]b[B]\cdots z[Z]) \uparrow T = a[AT]b[BT]\cdots z[ZT]$$

This pseudo-core algebra of trees is isomorphic to the free pseudo-core algebra on the given set of letters: the trees consisting of a single branch are the generators. In the more formal approach these are the sequences of the form $a[]$. (It is easily verified that the induced map from the free algebra to the algebra of trees is an isomorphism.)

We can construct free core algebras as quotients of free pseudo-core algebras. We need to formalize a few notions.

[13] The simplest such set of rules is that the brackets be "properly mated" and that every letter be immediately followed by a left-bracket, every left-bracket preceded by a letter. The left-brackets are thus revealed to be redundant. So an alternate definition of tree would be a sequence of letters and right-brackets such that the total number of letters and right-brackets are equal and for every initial subsequence the number of letters is at least as great as the number of right-brackets.

The number of branches from the root of a tree will be called its POD-NUMBER. (In the more formal approach: $a_1[A_1]a_2[A_2]\cdots a_n[A_n]$ has pod-number n.) A tree with pod-number 1 will be called a MONOPOD. A branch in an arbitrary tree gives rise to a SUBMONOPOD, to wit, the branch and the subtree sprouting therefrom.

Define a PRIMITIVE MODIFICATION on a tree as follows: given a pair of adjacent submonopods, a left one, A, and a right one, B, sprouting from a common node, first transpose them and then join a copy of B to the first node above the root of A by joining the root of the copy of B to the far right of all the branches sprouting from the first node above the root.

In the more formal approach a primitive modification is the rewrite rule that removes a subsequence of the form $a[A]b[B]$ and replaces it with $b[B]a[Ab[B]]$.

Define a binary relation on trees, $\mathcal{R}$, so that $(T)\mathcal{R}(T')$ means that T' is obtainable from T via a single primitive modification. The following are easily verified: If $(T)\mathcal{R}(T')$ then for any tree S it is the case that $(ST)\mathcal{R}(ST')$, $(TS)\mathcal{R}(T'S)$, $(T\uparrow S)\mathcal{R}(T'\uparrow S)$. What's missing is $(S\uparrow T)\mathcal{R}(S\uparrow T')$. It's a little more complicated: $(T)\mathcal{R}(T')$ implies $(S\uparrow T)\mathcal{R}^m(S\uparrow T')$ where m is the pod-number of S. All of which says that the equivalence relation defined by the existence of a sequence of primitive modifications and demodification that transforms one tree to another, that equivalence relation is a congruence (that is, if S and S' are equivalent then so are the pairs $\langle ST, S'T\rangle$, $\langle TS, TS'\rangle$, $\langle S\uparrow T, S'\uparrow T\rangle$ and $\langle T\uparrow S, T\uparrow S'\rangle$). Hence the set of equivalence classes has a unique algebra structure so that the function that assigns to each tree its equivalence class is a homomorphism.

The quotient algebra (that is, the set of equivalence classes) is a core algebra: equation 6 is satisfied because: $(ST)\mathcal{R}^n(T(S\uparrow T))$ where n is the product of the pod-numbers of S and T. This quotient algebra is easily verified to be the free core algebra.

Suppose, then, that $v\uparrow\phi = v\uparrow\phi'$. We may safely assume that ϕ and ϕ' are already in normal form. Necessarily there is a sequence of primitive modifications and demodifications transforming the tree that describes $v\uparrow\phi$ to the tree that describes $v\uparrow\phi'$. These trees are monopods. Any primitive modification (or demodification) requires a pair of adjacent branches from the same node, hence must take place above the bottom branch. Thus the same sequence of modifications and demodifications will transform the tree describing ϕ to the tree describing ϕ'. Done.

In the more formal approach base-cancellation becomes the observation that if there is a sequence of primitive modifications and demodifications that transforms $a[A]$ to $a[A']$ then the modifications and demodifications are necessarily taking place entirely between the brackets, hence the same sequence would transform A to A'.

The unrestricted base cancellation lemma reveals the complexity of the free core algebra on one generator. The free pseudo-core algebra may be identified with the set of (unlabeled) rooted planar trees. Equation 6 can not impose a bound on width (the free monoid on one generator, that is the natural numbers, appears as the set of rooted trees of height one) but before the unrestricted base cancellation lemma it seemed possible that equation 6 could make every such tree equal to one of bounded height. We now know that the terms $v, v\uparrow v, v\uparrow(v\uparrow v), v\uparrow(v\uparrow(v\uparrow v)), \ldots$ (those described by trees of width one) all name distinct elements.

Addenda

10. Proposition. *The number of 2-element core algebras (up to isomorphism) is 3.*

First, there are only 2 monoids of order 2, the 2-element group and the "Sierpinski" monoid ($\{1,0\}$ under multiplication). Any commutative monoid has the trivial core-algebra structure ($x \uparrow y = x$), any group has a unique core-algebra structure ($x \uparrow y = y^{-1}xy$), hence any abelian group has only the trivial structure. When the monoid is Sierpinski then equations $1, 1', 3, 3'$ leave only $0 \uparrow 0$ undefined; both possibilities work. The three core algebras are:

$$\begin{array}{c|cc} xy & 1 & 0 \\ \hline 1 & 1 & 0 \\ 0 & 0 & 1 \end{array} \qquad \begin{array}{c|cc} x\uparrow y & 1 & 0 \\ \hline 1 & 1 & 1 \\ 0 & 0 & 0 \end{array}$$

$$\begin{array}{c|cc} xy & 1 & 0 \\ \hline 1 & 1 & 0 \\ 0 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} x\uparrow y & 1 & 0 \\ \hline 1 & 1 & 1 \\ 0 & 0 & 0 \end{array}$$

$$\begin{array}{c|cc} xy & 1 & 0 \\ \hline 1 & 1 & 0 \\ 0 & 0 & 0 \end{array} \qquad \begin{array}{c|cc} x\uparrow y & 1 & 0 \\ \hline 1 & 1 & 1 \\ 0 & 0 & 1 \end{array}$$

11. Proposition. *The number of 3-element core algebras (up to isomorphism) is 16.*

First, there are exactly 7 monoid isomorphism types of order 3:

$$\begin{array}{ccc} 1 & a & b \\ a & b & 1 \\ b & 1 & a \end{array} \quad \begin{array}{ccc} 1 & v & 0 \\ v & 1 & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{ccc} 1 & e & v \\ e & e & v \\ v & v & e \end{array} \quad \begin{array}{ccc} 1 & v & 0 \\ v & 0 & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{ccc} 1 & v & 0 \\ v & v & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{ccc} 1 & a & b \\ a & a & b \\ b & a & b \end{array} \quad \begin{array}{ccc} 1 & a & b \\ a & a & a \\ b & b & b \end{array}$$

The proof that these are the only possibilities for 3-element monoids is as follows.

As in all finite monoids any element that has either a left or right inverse has both, that is, is a unit.[14] Hence if 1 appears in the multiplication table off the diagonal in any 3-element monoid it is necessarily a group, to wit, the 1st monoid above.

If 1 does not appear off the diagonal but does appear below the top row of the multiplication table (that is, if there's an involution) we obtain a sub-semigroup $\{1, v\}$ which is, in fact, a group. For any x, y such that $xy = v$ both x and y are units. The only way to avoid all three elements being units (in which case it would be have to be a 3-element group with an involution!) is for the remaining element to be an absorbing element, to wit, the 2nd monoid above

[14]Finiteness is unnecessary when there's a core-structure: $xy = 1$ implies $yx = yxxy = yxyx^y = yx^y = xy = 1$.

When 1 does not appear below the top line then the two elements different from 1 form a semigroup. We'll denote it S. The monoid is the result of formally adjoining a unit to S.

If S has an element that is not idempotent, say $v^2 \neq v$ then $S = \{v, v^2\}$. Having ruled out $v^3 = 1$ there are only two possibilities for v^3; they yield the 3rd and 4th monoids above.

We are left with the case that S is a 2-element idempotent semigroup. If it's commutative, that is, if S is a $\wedge$–semi-lattice, we obtain the 5th monoid above. A 2-element non-commutative idempotent semigroup satisfies either $xy = y$ or $xy = x$. The 6th monoid above is the case $xy = y$. The 7th monoid is its dual

The first five monoids are commutative and therefore have the trivial core-algebra structure ($x \uparrow y = x$).

We will often use that $(-)\uparrow y$ is a core-algebra endomorphism, in particular, that if x is an idempotent, unit, or involution, then so is $x \uparrow y$.

The 1st monoid, being a group, can have only one core-algebra structure ($x \uparrow y = y^{-1}xy$) which in this case is, of course, the trivial.

xy	1	a	b
1	1	a	b
a	a	b	1
b	b	1	a

$x\uparrow y$	1	a	b
1	1	1	1
a	a	a	a
b	b	b	b

The 2nd monoid has three core-algebra structures. From $xv = vx^v$ we may cancel to obtain $x^v = x$. From $(v^0)^2 = (vv)^0 = 1^0 = 1$ we know that v^0 is either 1 or v. If it's v then $0^0 v = 0^0 v^0 = 0^0$ forcing $0^0 = 0$ which yields the trivial core-structure. If, instead, $v^0 = 1$ there are two idempotent possibilities for 0^0 and both work. The three core structures:

xy	1	v	0
1	1	v	0
v	v	1	0
0	0	0	0

$x\uparrow y$	1	v	0
1	1	1	1
v	v	v	v
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	1	0
0	0	0	0

$x\uparrow y$	1	v	0
1	1	1	1
v	v	v	1
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	1	0
0	0	0	0

$x\uparrow y$	1	v	0
1	1	1	1
v	v	v	1
0	0	0	1

The 3rd monoid has only the trivial core-structure.

xy	1	e	v
1	1	e	v
e	e	e	v
v	v	v	e

$x\uparrow y$	1	e	v
1	1	1	1
e	e	e	e
v	v	v	v

The proof is as follows: from $e = vv = v(v^v)$ we know that $v^v \neq 1$. Since there's only one square root of 1 we further know $(v^v)^2 \neq 1$ hence $e^v = (vv)^v = (v^v)^2 \neq 1$. Since e^v is idempotent we know that it must be the other idempotent, hence $e^v = e$. And $e^e \neq 1$ else $e = e^v = e^{ev} = (e^e)^v = 1^v = 1$ hence e^e must be the other idempotent, that is, $e^e = e$.

$v^e = 1$ leads to a contradiction as follows: $e = vv = v(v^v) = v(v^{ev}) = v(v^e)^v = v(1^v) = v1 = v$. Since, therefore, $v^e \in S$ we may cancel e from $ve = e(v^e)$ to obtain $v = v^e$.

From $vv = v(v^v)$ we may exclude v^v being either e or 1. Hence $v^v = v$.

For the 4th monoid there are two core-structures.

xy	1	v	0
1	1	v	0
v	v	0	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	v
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	0	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	0	0
0	0	0	0

First, $0 = 0v = v(0^v)$ prohibits $0^v = 1$. Since 0^v must be idempotent, we conclude that $0^v = 0$.

v^v can not be 1 (because then $1 = (v^v)^2 = (vv)^v = 0^v = 0$).

If $v^v = v$ then $v^0 = v^{vv} = (v^v)^v = v^v = v$ and $0^0 = (vv)^0 = (v^0)^2 = v^2 = 0$. This is the trivial core structure.

If $v^v = 0$ then $v^0 = v^{vv} = (v^v)^v = 0^v = 0$ and $0^0 = (vv)^0 = (v^0)^2 = 0^2 = 0$. This is the core-structure that appeared as the second specialization in the proof of the Critical Lemma.

The 5th monoid requires a little work. We view $\{1, v, 0\}$ as a linearly ordered set with xy the smaller of x, y. Equation 5 for core algebras says that x^y is order-preserving for fixed y. Equation 6, $xy = yx^y$, has no consequence when y is either 1 or 0 but $0v = v0^v$ forces $0^v = 0$ and $vv = vv^v$ prohibits $v^v = 0$. There are two consequences of equation 4: if $v^v = 1$ then $v^0 = v^{v0} = (v^v)^0 = 1^0 = 1$ and if $0^0 = v$ then $v^0 = (0^0)^0 = 0^{00} = 0^0 = v$. These are the only restrictions. The seven structures in lexicographic order are:

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	0
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	v
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	v
0	0	0	v

xy	1	v	0
1	1	v	1
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	1
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	v	1
0	0	0	1

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	1	1
0	0	0	0

xy	1	v	0
1	1	v	0
v	v	v	0
0	0	0	0

$x{\uparrow}y$	1	v	0
1	1	1	1
v	v	1	1
0	0	0	1

The 6$^{\text{th}}$ monoid has two core structures. First note that if $x^y = 1$ for any $x, y \in S$ then for all $u, w \in S$ we have $u^w = (ux^y)^w = u^w x^{yw} = u^w x^w = (ux)^w = x^y = 1$ which yields:

xy	1	a	b
1	1	a	b
a	a	a	b
b	b	a	b

$x{\uparrow}y$	1	a	b
1	1	1	1
a	a	1	1
b	b	1	1

If, on the other hand, $x^y \neq 1$ for all $x, y \in S$ then $xy = yx^y$ forces $y = x^y$. We obtain the core algebra obtained from the 2-element semi-core algebra in which both binary operations satisfy $x \star y = y$:

xy	1	a	b
1	1	a	b
a	a	a	b
b	b	a	b

$x{\uparrow}y$	1	a	b
1	1	1	1
a	a	a	b
b	b	a	b

The 7$^{\text{th}}$ monoid is the only one that has no core-algebra structure (if it did, then $a = ab = b(a^b) = b$).

The numbers of core-structures for the seven monoids of order 3 are thus:

1	2	3	4	5	6	7
1	3	1	2	7	2	0

Dept. of Mathematics, University of Pennsylvania
Philadelphia, PA 19104
Email: pjf@cis.upenn.edu

This article may be accessed at http://www.tac.mta.ca/tac/ or by anonymous ftp at ftp://ftp.tac.mta.ca/pub/tac/html/volumes/18/11/18-11.{dvi,ps,pdf}