UNIVERSITY OF TASMANIA

DEPARTMENT OF MATHEMATICS

MINOR CLASSES

by

Dirk Llewellyn Vertigan B.Sc. (Hons)

Submitted in fulfilment of the requirements for

the degree of

Master of Science

August 1988

Except as stated herein, this thesis contains no material which has been accepted for the award of any other degree or diploma in any university, and to the best of my knowledge and belief, contains no copy or paraphrase of material previously published or written by another person except where duly acknowledged.

D. L. Vertigan

Dirk Vertigan

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Don Row and also Dr. Geoff Whittle for his guidance, encouragement and unending patience.

My sincere gratitude goes to Mrs. Betty Golding for her excellent typing under the most hurried conditions, working hours far beyond the demands of duty.

Abstract: For many classes of combinatorial structures, such as graphs and matroids, there exists a concept of point removal. Minors are obtained by various manners of point removal. In this thesis, these ideas are abstracted to give the definition of minor class. It turns out that minor classes are algebras, in the sense of universal algebra, which makes much universal algebra theory available to the study of minor classes. For example, varieties of minor classes are studied, as well as subalgebras (sub minor classes), homomorphisms, and direct products. Amongst the theory developed, is a natural connection between varieties of minor classes and categories. Also it is shown how a minor class can be described in terms of its so called ψ-structures and natural excluded minors (which are its excluded minors in the so called completion of the minor class). Many well known minor classes are described in this way, including the minor class of matroids, various minor classes of graphs, and minor classes of subspaces of certain vector spaces over a field (related to Tutte's chain groups). For any field, the latter minor class has, as a homomorphic image, the minor class of matroids coordinatisable over that field. This provides a motivation for further study of minor class homomorphisms.

CONTENTS

SECTION 1: INTRODUCTION

This paper draws from a range of mathematical topics, but no specialist knowledge is assumed. Of course, it is unavoidable that the introduction mentions terms which are not defined until later in the text. Section 2 explains some notation and conventions used in this paper.

In the literature there are many examples of *structures* based on their *ground sets.* Isomorphisms between structures are induced in a natural way by bijections between their ground sets. In some cases, for example matroids [18] and graphs [16], a structure can have *points* (ground set elements) *removed* in various *manners*, yielding another structure whose ground set lacks the removed points. This structure is a *minor* of the original structure. A *minor class* consists of a collection of structures together with the operations of point removal and isomorphism, provided certain desirable conditions are satisfied. Section 3 presents these ideas formally. Thus minor classes abstract certain aspects common to various topics in the literature, and the study of minor classes has consequences for these topics.

Minor classes of a certain form are presented as an example in section 4. An enlightening visualisation is given for their structures and for the operations of point removal and isomorphism. In section 8 it is shown that any minor class can be *embedded* (see section 5) in a minor class of the form described in section 4, so that the given visualisation, can be used in general.

Section 5 shows that minor classes are algebras, in the sense of universal algebra [1,6,5,3], which immediately makes a large-body of universal algebra theory available to the study of minor classes. The elements of the algebra are the structures, and the operations of the algebra are the operations of point removal and isomorphism. The conditions imposed on minor classes are of a form central to universal algebra, namely *equations.* Thus we can talk of *varieties* of minor classes, as well as *sub minor classes*, minor class *homomorphisms*, and *direct products* of minor classes. Section 6 presents some minor class constructions related to more peripheral parts of universal algebra theory.

Birkhoff's equivalent characterisations of varieties are presented in section 7, together with similar characterisations for *proper varieties* and *regular unary varieties.* (Varieties of minor classes are proper and regular unary varieties.) *Free algebras,* in particular free minor classes, are defined. Section 9 shows how a *category* can be associated with a regular unary variety, and describes this category in the minor class case. There is an intimate connection between the category, and the free algebras, freely generated by one element. When the category or the free algebras satisfy certain conditions, the regular unary variety is a *special unary variety* (which includes the minor class case).

Complete algebras are defined in section 10, and in a special unary variety the *completion* of an algebra is defined, which is the "smallest" complete algebra having the given algebra as a subalgebra. Other results of an algebraic nature are given.

The theory of section 10 is specialised to minor classes (in which ground sets are finite, the case of current interest) in section 11. It is shown how a minor class can be described in terms of its so-called *ψ-structures* and *natural excluded minors* (which are its excluded minors in the completion of the minor class). Many well known minor classes are described in this way in sections 12, 13 and 14, and this description is often surprisingly simple. For example, the minor class of *matroids* [18] has only two ψ-structures and six natural excluded minors.

Section 12 examines the minor class of matroids, which is a sub minor class of the minor class of closure operators. Section 13 examines minor classes of linear *dependencies* which have as homomorphic images, minor classes of matroids coordinatisable over a field. It becomes apparent that the study of minor class homomorphisms could have consequences for some difficult combinatorial problems. Minor classes of *graphs* and *digraphs* [16], in which the ground sets are either the edge sets or the vertex sets, are examined in section 14, together with some interesting homomorphisms.

The paper concludes with section 15 presenting some ideas for further research, as well as some conjectures and counterexamples.

SECTION 2: NOTATION AND CONVENTIONS

Let A and B be sets. The power set of A is denoted 2^A and the cardinality of A is denoted $|A|$. The set difference A–B denotes $\{a|a\in A$ and $a\notin B\}$.

Let I be a set. A function $f:I\to A$ can be denoted as a *family* or *vector* $(f(i)|i\in I)$. If $(A^i|i\in I)$ is a family of sets, then its cartesian product $\prod_{i\in I} A^i$ is the set $\{(a_i|i\in I)|a_i\in A^i$ for every $i\in I\}$. In particular $A^I= \prod_{i\in I} A$ is the set of all functions from A to I.

Let $\mathcal{Q}$ be a set. A *$\mathcal{Q}$-labelled partition*, or a *$\mathcal{Q}$-partition,* of a set A is given by a function $f:A\to\mathcal{Q}$ (associating a "label" $f(a)\in\mathcal{Q}$ with each $a\in A$) and is denoted $<A_s|s\in\mathcal{Q}>$ where $A_s=\{a|a\in A$ and $f(a)=s\}$ for each $s\in\mathcal{Q}$. It is convenient to write $A=<A_s|s\in\mathcal{Q}>$ to indicate that A can be treated as a set, or when relevant, can be partitioned into disjoint subsets labelled by elements of $\mathcal{Q}$.

Let $B=<B_s|s\in\mathcal{Q}>$ and $C=<C_s|s\in\mathcal{Q}>$. Define $B\subseteq C$ to mean $B_s\subseteq C_s$ for every $s\in\mathcal{Q}$. Define $B\supseteq C$ and $B=C$ similarly. Let $A=<A_s|s\in\mathcal{Q}>$ and suppose $A^i=<A^i_s|s\in\mathcal{Q}>$ and $A^i\subseteq A$ for every $i\in I$. Define $\bigcap_{i\in I} A^i$ to be $<\bigcap_{i\in I} A^i_s|s\in\mathcal{Q}>$ and define $\bigcup_{i\in I} A^i$ to be $<\bigcup_{i\in I} A^i_s|s\in\mathcal{Q}>$. Define $\prod_{i\in I} <A^i_s|s\in\mathcal{Q}>$ to be $<\prod_{i\in I} A^i_s|s\in\mathcal{Q}>$. (The notation $\prod_{i\in I} A^i$ risks being ambiguous.)

Let $A=<A_s|s\in\mathcal{Q}>$ and $B=<B_s|s\in\mathcal{Q}>$. A function $\alpha:A\to B$ is *$\mathcal{Q}$-respecting* if $\alpha(a)\in B_s$ for every $s\in\mathcal{Q}$ and every $a\in A_s$. In this case we can talk of the $\mathcal{Q}$-partition $<\alpha_s:A_s\to B_s|s\in\mathcal{Q}>$ of $\alpha:A\to B$, where $\alpha_s(a)=\alpha(a)$ for every $s\in\mathcal{Q}$ and every $a\in A_s$. This situation is expressed as $(\alpha:A\to B)=<\alpha_s:A_s\to B_s|s\in\mathcal{Q}>$.

Let $A=<A_s|s\in\mathcal{Q}>$. A relation $q\subseteq A\times A$ is *$\mathcal{Q}$-respecting* if $(a,b)\in q$ (denoted aqb) implies that $a,b\in A_s$ for some $s\in\mathcal{Q}$. In this case we can talk of the $\mathcal{S}$-partition $<q_s \subseteq A_s \times A_s | s \in \mathcal{S}>$ of $q \subseteq A \times A$ where $a q_s b$ exactly when $a, b \in A_s$ and aqb.

A *complete lattice* can be thought of as a partially ordered set with the property that for any set of elements, there is a unique smallest element greater than the elements, called their *join*, and a unique greatest element smaller than the elements, called their *meet*. (Often join and meet are defined first and the partial order derived from them.)

Often a symbol calls for a set to be written in a certain position, for example the set X in $T^{\Sigma}(X)/eq^{X}(\mathscr{E})$. But if X is the single element set $\{x\}$ then it is more convenient to replace $\{x\}$ by x in the symbol, for example $T^{\Sigma}(x)/eq^{x}(\mathscr{E})$ rather than $T^{\Sigma}(\{x\})/eq^{\{x\}}(\mathscr{E})$.

Most papers do not precisely define "set" and neither will this one. However, it is assumed that "set" is well defined, and each set is a well defined collection of well defined elements (although not all such collections can be sets, as Russell's Paradox shows). It is also assumed that the definition of "set" satisfies certain desirable criteria (such as: any subset of a set is a set, the power set of a set is a set, the union and cartesian product of a family of sets are sets). Let "class" and "metaclass" be defined subject to the same criteria, but with every set (respectively class) being a class (respectively metaclass) and the collection of all sets (respectively classes) being a class (respectively metaclass), so that the definitions themselves must be different. Thus, a theorem using the words "set" and "class" (but not "metaclass") which does not make use of the particular definition of these words, could have "set" (respectively "class") changed to "class" (respectively "metaclass") and remain true. The modified theorem would appear more general, relative to a fixed definition of "set" and "class" but really it would be saying exactly the same thing as the original theorem.

SECTION 3: MINOR CLASS DEFINITIONS

This section develops the definition of minor class in stages. The elements of a minor class are called structures, and associated with each structure is a set called its ground set. Isomorphisms between structures are represented by bijections between their ground sets. Finally there are operations of point removal, which when applied to one structure, yield another structure whose ground set is obtained by removing points from the ground set of the first structure.

A *structure* is a pair (S,Q) where Q is a set. (If the structure is a well known mathematical object, its conventional name is used.) When Q is known from the context, (S,Q) is abbreviated to S. The set Q is the *ground set of* S and is denoted by G(S). While no conditions are imposed on S in the pair (S,Q), each concrete example of a structure in this paper satisfies the intuitive notion of being based on its ground set, in the way that a matroid or group or ring is based on a set.

A *structure class* consists of a class $\mathscr{S}$ of structures, together with a class $\mathscr{Q}$ of sets, such that $G(S)\in\mathscr{Q}$ for every structure $S\in\mathscr{S}$. The elements of $\mathscr{Q}$ are *ground sets.* This structure class can be denoted by the pair $(\mathscr{S},\mathscr{Q})$ or abbreviated to $\mathscr{S}$ when $\mathscr{Q}$ is clear from the context. Often the symbol $\mathscr{Q}$ is not mentioned, but rather, the sets which are ground sets are specified. For the structure class $(\mathscr{S},\mathscr{Q})$ and any ground set $Q\in\mathscr{Q}$, let $\mathscr{S}_Q$ denote $\{S|S\in\mathscr{S},G(S)=Q\}$, that is, the class of structures in $\mathscr{S}$ which have ground set Q. For distinct ground sets $P,Q\in\mathscr{Q}$ it is clear that $\mathscr{S}_P\cap\mathscr{S}_Q$ is empty and hence $<\mathscr{S}_Q|Q\in\mathscr{Q}>$ is a $\mathscr{Q}$-partition of $\mathscr{S}$. It is permissible that $\mathscr{S}_Q$ be empty for some ground set $Q\in\mathscr{Q}$ (so that there are no structures in $\mathscr{S}$, on ground set Q) because when various structure classes are compared, it is desirable that they have the same class of ground sets.

Consider the following example for which the ground sets are exactly the finite sets. Let $\mathscr{W}$ be the class of all pairs (W,Q) where Q is a finite set and $W \subseteq 2^Q$. (Observe that W satisfies the intuitive notion of a structure based on the set Q.) Clearly $\mathscr{W}$ is a structure class with finite ground sets.

If a structure S is genuinely based on its ground set Q, then for any ground set $P \in \mathscr{Q}$, a bijection $\omega: Q \to P$ induces an isomorphic structure on ground set P in a natural way. A description of the isomorphic structure naturally induced by ω is obtained from a description of the structure S, by replacing every occurrence of each $q \in Q$ (in the description of S) by $\omega(q)$. (For example, consider this construction applied to the multiplication table of a group whose elements are the elements of Q.) However, it is desirable to axiomatise the behaviour of isomorphism for the more abstract definition of structure given above, as the following definition does.

An *isomorphism class* consists of a structure class $(\mathscr{S}, \mathscr{Q})$ together with a class $\mathscr{I}$ of functions; one function from $\mathscr{S}_Q$ to $\mathscr{S}_P$ for all ground sets $Q, P \in \mathscr{Q}$ and every bijection $\omega: Q \to P$. This function maps each structure $S \in \mathscr{S}_Q$ to a structure in $\mathscr{S}_P$ denoted $\omega(S)$, *the isomorphic copy of* S *induced by* the bijection $\omega: Q \to P$. (It may seem that $\omega(S)$ is an abuse of notation, since in general $S \not\subseteq Q$, but there is no ambiguity and the notation is convenient.) Also, the structure $\omega(S)$ is (*structure*) *isomorphic* to the structure S, via the (*structure*) *isomorphism* ω. If the structure $T \in \mathscr{S}_P$ is such that $T = \omega(S)$ then $T \cong S$. An isomorphism class is required to satisfy the following two conditions.

(M1) For any ground set $Q \in \mathscr{Q}$, if $1: Q \to Q$ is the identity function, then $1(S) = S$ for every structure $S \in \mathscr{S}_Q$.

(M2) For all ground sets $Q, P, R \in \mathscr{Q}$ and all bijections $\omega: Q \to P$ and $\tau: P \to R$ it holds that $\tau(\omega(S)) = (\tau \circ \omega)(S)$, for every structure $S \in \mathscr{S}_Q$.

It is clear that (M1) and (M2) will hold if all structures are genuinely based on their ground sets, and isomorphism is defined "naturally" as described above. All the concrete examples in this paper define isomorphism naturally.

For any ground set $Q\in\mathscr{Q}$ and any structure $S\in\mathscr{S}_Q$, a (*structure*) *automorphism* of S is a (structure) isomorphism from S to itself, that is, a bijection $\omega{:}Q\to Q$ (a *permutation* of Q) such that $\omega(S)=S$. Let Aut(S) denote $\{\omega\,|\,\omega{:}Q\to Q$ is a bijection and $\omega(S)=S\}$, the set of (structure) automorphisms of S. Conditions (M1) and (M2) ensure that Aut(S) is a group of permutations of Q, with respect to composition of functions. For any bijection $\tau{:}Q\to P$, the automorphisms of structure $S\in\mathscr{S}_Q$ uniquely determine the automorphisms of structure $\tau(S)\in\mathscr{S}_P$ since if ω is a permutation of Q, then $\omega(S)=S$ if and only if $(\tau\circ\omega\circ\tau^{-1})(\tau(S))=\tau(S)$. (This uses conditions (M1) and (M2). Observe that $\tau\circ\omega\circ\tau^{-1}$ is a permutation of P.) Thus $\mathrm{Aut}(\tau(S))=\{\tau\circ\omega\circ\tau^{-1}\,|\,\omega\in\mathrm{Aut}(S)\}$.

Returning to our previous example, the structure class $\mathscr{W}$ becomes an isomorphism class if isomorphism is defined in the natural way, as follows. For ground sets $Q,P\in\mathscr{Q}$, let $\omega{:}Q\to P$ be a bijection. Any structure $W\in\mathscr{W}_Q$ can be explicitly written out as a set of subsets of Q, and so replacing every occurrence of each $q\in Q$ by $\omega(q)$, yields $\omega(W)$ explicitly written out as a set of subsets of P. More formally, define $\omega{:}2^Q\to 2^P$ by setting $\omega(N)=\{\omega(q)\,|\,q\in N\}$ for every $N\subseteq Q$, and define $\omega{:}2^{(2^Q)}\to 2^{(2^P)}$ similarly. (Recall that $(W,Q)\in\mathscr{W}$, exactly when $W\subseteq 2^Q$, that is, $W\in 2^{(2^Q)}$.) So the isomorphism ω sends (W,Q) to $(\omega(W),P)$ as required. Conditions (M1) and (M2) are guaranteed to hold when isomorphism is defined "naturally", so that $\mathscr{W}$ is indeed an isomorphism class.

The elements of the ground set of a structure are *points*, except in specific cases where *vertices* or *edges* are more appropriate names. In the mathematical literature, there are many examples of structures (for example matroids [18] or graphs [16]) which can have points "removed" from them in various manners, yielding another structure whose ground set lacks the removed points. The definitions below develop this idea.

A class of sets $\mathscr{Q}$ is *hereditary* if every subset of each element of $\mathscr{Q}$, also is an element of $\mathscr{Q}$. That is, if $Q \in \mathscr{Q}$ and $P \subseteq Q$, then $P \in \mathscr{Q}$. A *pre point–removal class* consists of an isomorphism class $(\mathscr{S}, \mathscr{Q}, \mathscr{I})$ where $\mathscr{Q}$ is a hereditary class of finite sets, together with a class K and another class $\mathscr{P}$ of functions; one function from $\mathscr{S}_Q$ to $\mathscr{S}_{Q-\{q\}}$ for every ground set $Q \in \mathscr{Q}$, every point $q \in Q$ and every *manner* $\ell \in K$ (as the elements of K are called). This function sends each structure $S \in \mathscr{S}_Q$ to a structure in $\mathscr{S}_{Q-\{q\}}$ denoted $S[\ell,q]$. The structure $S[\ell,q]$ is *obtained from* S *by removing point* q *in manner* ℓ. In addition to (M1) and (M2), a pre point–removal class is required to satisfy the following condition.

(M3') For all ground sets $Q, P \in \mathscr{Q}$, every bijection $\omega: Q \to P$, every point $q \in Q$ and every manner $\ell \in K$, it holds that

$\omega|_{Q-\{q\}}(S[\ell,q]) = (\omega(S))[\ell,\omega(q)]$ for every structure $S \in \mathscr{S}_Q$.

The bijection $\omega|_{Q-\{q\}}: Q-\{q\} \to P-\{\omega(q)\}$ is the restriction of ω to $Q-\{q\}$ and is used because $S[\ell,q]$ has ground set $Q-\{q\}$. Condition (M3') simply ensures that point removal behaves sensibly with respect to isomorphism. This is guaranteed if isomorphism is defined naturally (so that (M1) and (M2) hold) and the definition of point removal is invariant with respect to renaming ground set elements. This applies to all the concrete examples in this paper, so that (M3'), along with (M1) and (M2), are guaranteed to hold.

Returning again to our example, one way to make $\mathscr{W}$ into a pre point–removal class is as follows. Let K={delete, contract}. For every ground set (finite set) Q, every structure $W \in \mathscr{W}_Q$ and every point $q \in Q$, let $W\backslash q$ denote W[delete,Q] and let W/q denote W[contract,q]. Define $W\backslash q$, the *deletion* of q from W, to be $\{w \mid w \in W \text{ and } q \notin w\}$ and define W/q, the *contraction* of q from W, to be $\{w-\{q\} \mid w \in W \text{ and } q \in w\}$. Observe that $W\backslash q$ and W/q are both subsets of $2^{Q-\{q\}}$ and hence elements of $\mathscr{W}_{Q-\{q\}}$ as required. By earlier comments, condition (M3') clearly holds, so that $\mathscr{W}$ is indeed a pre point–removal class.

The examples of point removal in the mathematical literature generally have the property that points can be removed "in any order" yielding the same result. The following definition incorporates this property.

A *point–removal class* is a pre point–removal class in which the following condition holds.

(M4') For every ground set $Q\in\mathscr{Q}$, all <u>distinct</u> points $q,r\in Q$ and all manners $\ell,m\in K$ it holds that

$(S[\ell,q])[m,r]=(S[m,r])[\ell,q]$ for every structure $S\in\mathscr{S}_Q$

(It is necessary that q and r be distinct, since once a point is removed, it is absent from the ground set and cannot be removed again. However ℓ and m need not be distinct.) Condition (M4') says that, starting with a structure $S\in\mathscr{S}_Q$, the effect of removing point q in manner ℓ and then removing point r in manner m, is the same as first removing point r in manner m and then point q in manner ℓ. It follows by induction that a sequence of point removals can be performed in any order yielding the same result (provided the manner of removing any particular point is not changed).

Our example, the pre point–removal class $\mathscr{W}$, is in fact a point–removal class, since for any ground set (finite set) Q, any structure $W\in\mathscr{W}_Q$, and all distinct points $q,r\in Q$, it holds that

$$(W\backslash r)\backslash q=\{w\,|\,w\in W \text{ and } q,r\notin w\}=(W\backslash q)\backslash r\ ,$$
$$(W\backslash r)/q=\{w-\{q\}\,|\,w\in W \text{ and } q\in w \text{ and } r\notin w\}=(W/q)\backslash r\ ,$$
$$(W/r)/q=\{w-\{q,r\}\,|\,w\in W \text{ and } q,r\in w\}=(W/q)/r\ ,$$

as required.

Given a point–removal class $\mathscr{S}$ and a ground set Q, it is instructive to determine all the possibilities for removing some (or no) points from the structures in $\mathscr{S}_Q$. Since the order of the point removals makes no difference (as guaranteed by condition (M4')), it suffices to specify for each point $q \in Q$, whether or not q is removed, and if it is, in which manner (an element of K) it is removed. This can be encapsulated in a single symbol as follows. Let $\overline{K}=K \cup \{\odot\}$ where $\odot \notin K$. (This definition remains in place throughout the paper.) For any *prescription* $\mathfrak{K} \in \overline{K}^Q$ (that is, $\mathfrak{K}:Q \to \overline{K}$), let $G(\mathfrak{K})=\{q \mid q \in Q$ and $\mathfrak{K}(q)=\odot\}$ and let $S[\mathfrak{K}]$ be the structure with ground set $G(\mathfrak{K})$, obtained from the structure $S \in \mathscr{S}_Q$ by removing point q in manner $\mathfrak{K}(q)$ for each point $q \in Q-G(\mathfrak{K})$, and not removing any point in $G(\mathfrak{K})$. It does not matter in which order the points are removed, and in fact they could be considered to be removed simultaneously.

So far, in this discussion about point removals, all ground sets have been finite. One could allow infinite ground sets, but only finitely many points can be removed by removing only one point at a time. However, in the definition of $S[\mathfrak{K}]$ in the previous paragraph, one can consider all points in $Q-G(\mathfrak{K})$ to be removed in a single operation determined by the prescription $\mathfrak{K}$. So it is reasonable to allow $Q-G(\mathfrak{K})$ to be infinite when Q is, although $S[\mathfrak{K}]$ could not then be obtained from S by removing one point at a time. A new formulation is required, and is given below.

A *minor class* consists of an isomorphism class $(\mathscr{S}, \mathscr{Q}, \mathscr{I})$ where $\mathscr{Q}$ is a hereditary class of sets, together with a class K and another class $\mathscr{P}$ of functions; one function from $\mathscr{S}_Q$ to $\mathscr{S}_{G(\mathfrak{K})}$ for every ground set $Q \in \mathscr{Q}$ and every prescription $\mathfrak{K} \in \overline{K}^Q$. This function sends each structure $S \in \mathscr{S}_Q$ to a structure in $\mathscr{S}_{G(\mathfrak{K})}$ denoted $S[\mathfrak{K}]$. In addition to (M1) and (M2), a minor class is required to satisfy the three conditions (M3), (M4) and (M5), given below.

Firstly, two definitions are needed. For all ground sets $Q,P \in \mathscr{Q}$, any prescription $\mathfrak{K} \in \overline{K}^Q$ and any bijection $\omega:Q \to P$, the prescription $\omega(\mathfrak{K}) \in \overline{K}^Q$ is given by $(\omega(\mathfrak{K}))(\omega(q))=\mathfrak{K}(q)$ for all $q\in Q$. (That is, $\omega(\mathfrak{K})$ is actually $\mathfrak{K}\circ\omega^{-1}$.) For all classes A,B,C,D with $B\subseteq A$, and all functions $f:A\to C$ and $g:B\to D$, the function $(f\Delta g):A\to C\cup D$ is given by setting $(f\Delta g)(b)=g(b)$ whenever $b\in B$ and $(f\Delta g)(a)=f(a)$ whenever $a\in A-B$. In particular, for prescriptions $\mathfrak{J}\in\overline{K}^{Q}$ and $\mathfrak{L}\in\overline{K}^{G(\mathfrak{J})}$, it follows that the prescription $\mathfrak{K}\in\overline{K}^{Q}$, where $\mathfrak{K}=\mathfrak{J}\Delta\mathfrak{L}$ is such that $\mathfrak{K}(q)=\mathfrak{L}(q)$ whenever $q\in G(\mathfrak{J})$ and $\mathfrak{K}(q)=\mathfrak{J}(q)$ whenever $q\in Q-G(\mathfrak{J})$.

The three conditions are as follows.

(M3) For all ground sets $Q,P\in\mathcal{Q}$, every prescription $\mathfrak{K}\in\overline{K}^{Q}$ and every bijection $\omega:Q\to P$ it holds that

$\omega|_{G(\mathfrak{K})}(S[\mathfrak{K}])=(\omega(S))[\omega(\mathfrak{K})]$ for every structure $S\in\mathscr{S}_Q$.

The bijection $\omega|_{G(\mathfrak{K})}$ is the restriction of ω to $G(\mathfrak{K})$ and is used because $S[\mathfrak{K}]$ has ground set $G(\mathfrak{K})$. Condition (M3), like (M3'), simply ensures that point removal behaves sensibly with respect to isomorphism.

(M4) For every ground set $Q\in\mathcal{Q}$, if the prescription $\mathfrak{N}\in\overline{K}^{Q}$ is such that $G(\mathfrak{N})=Q$ (that is, $\mathfrak{N}(q)=\odot$ for every point $q\in Q$) then $S[\mathfrak{N}]=S$ for every structure $S\in\mathscr{S}_Q$.

This says that removing no points from S leaves it unchanged.

(M5) For every ground set $Q\in\mathcal{Q}$ and all prescriptions $\mathfrak{J}\in\overline{K}^{Q}$ and $\mathfrak{L}\in\overline{K}^{G(\mathfrak{J})}$, if prescription $\mathfrak{K}=(\mathfrak{J}\Delta\mathfrak{L})\in\overline{K}^{Q}$, then $S[\mathfrak{K}]=(S[\mathfrak{J}])[\mathfrak{L}]$ for every structure $S\in\mathscr{S}_Q$.

This says that the effect of removing some points, prescribed by $\mathfrak{J}$, and then removing some more points, prescribed by $\mathfrak{L}$, is the same as removing them all at once, as prescribed by $\mathfrak{K}=\mathfrak{J}\Delta\mathfrak{L}$. (Note that each point $q\in Q-G(\mathfrak{K})$ is removed in manner $\mathfrak{K}(q)$ in both cases.)

For any structure $S\in\mathscr{S}_Q$ and any prescription $\mathfrak{K}\in\overline{K}^{Q}$, the structure $S[\mathfrak{K}]$ is a *minor* of S. If $G(\mathfrak{K})\neq G(S)$ (that is more than zero points are removed), then $S[\mathfrak{K}]$ is a *proper minor* of S. A structure which is isomorphic to a minor of S is an *isominor* of S.

In the case that all ground sets are finite, the definitions of point–removal class and minor class are equivalent (up to change of notation), since all the mappings $S \to S[\ell,q]$ determine all the mappings $S \to S[\mathfrak{K}]$ and visa versa, and the conditions on these mappings (and the structure isomorphisms) are equivalent. This means that one can use either formulation, or even mix the two, whatever is most convenient. (Also, when necessary, one can verify whichever conditions are easier to show.)

A minor class can be denoted by the quintuple $(\mathscr{S},\mathscr{Q},\mathscr{I},K,\mathscr{P})$ often abbreviated to $\mathscr{S}$. Generally $\mathscr{Q}$, $\mathscr{I}$ and $\mathscr{P}$ (and quite often also K) remain unnamed, while their elements are all specified. For an alleged minor class $(\mathscr{S},\mathscr{Q}\ \mathscr{I},K,\mathscr{P})$ it is necessary to verify that conditions (M1)–(M5) hold. As discussed earlier, conditions (M1)–(M3) are guaranteed when isomorphism is defined naturally and is respected by point removals, a case that is easily recognised. Conditions (M4) and (M5) should be checked, although (M4) is generally trivial. Note that the conditions (M1)–(M5) are independent, that is, it is possible to construct a quintuple $(\mathscr{S},\mathscr{Q},\mathscr{I},K,\mathscr{P})$ satisfying any subset of these conditions, and only that subset.

Recall our example of a point–removal class $\mathscr{W}$. It follows that $\mathscr{W}$ is a minor class. The idea of $\mathscr{W}$ is easily generalised to a minor class $\mathscr{W}^{\mathscr{Q}}$ (for any hereditary class of sets $\mathscr{Q}$) with $\mathscr{Q}$ being the class of ground sets and $K = \{\text{delete, contract}\}$. Adopting a convenient notation let $W|^{B}_{A}=\{w-B \mid w\in W$ and $A\cap w=\emptyset$ and $B\subseteq w\}\subseteq 2^{Q-(A\cup B)}$ be the structure on ground set $Q-(A\cup B)$ obtained from (W,Q) by deleting all points in A and contracting all points in B, where $A,B\subseteq Q$ and $A\cap B=\emptyset$. Defining isomorphism naturally (as done earlier for $\mathscr{W}$) conditions (M1), (M2) and (M3) hold. Obviously $W|^{\emptyset}_{\emptyset}=W$ and for any disjoint subsets A,B,C,D of Q, both $(W|^{B}_{A})|^{D}_{C}$ and $W|^{(B\cup D)}_{(A\cup C)}$ are equal to $\{w-(B\cup D) \mid w\in W$ and $(A\cup C)\cap w=\emptyset$ and $(B\cup D)\subseteq w\}$. Therefore $\mathscr{W}^{\mathscr{Q}}$ is indeed a minor class.

SECTION 4: EXAMPLES OF MINOR CLASSES

Let K,C,B be classes with $K\cap C=\emptyset$, and let $A=K\cup C$, for the duration of this section. Let $\mathcal{Q}$ be a hereditary class of sets. In this section, a minor class, denoted $\mathcal{F}^{\mathcal{Q}}(K,C,B)$, is defined, where $\mathcal{Q}$ is the class of ground sets and K is the class of manners (of point removal). This is quite a general example because any minor class can, in a sense to be defined later, be "embedded" in a minor class of the form $\mathcal{F}^{\mathcal{Q}}(K,C,B)$.

The definition is developed in stages, starting with $\mathcal{F}^{\mathcal{Q}}(K,C,B)$ as a structure class. The structures in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$ with ground set $Q\in\mathcal{Q}$ are all the pairs (f,Q) where $f\in B^{(A^Q)}$, that is, f is a function from A^Q to B. Note that the elements of A^Q are themselves functions from Q to A, and can be denoted as vectors (section section 2) $x=(x_q|q\in Q)$ or $(x(q)|q\in Q)$ where $x_q=x(q)\in A$ for all $q\in Q$.

These structures, that is, functions of the form $f{:}A^Q\to B$, can be conveniently visualised in the special case that Q,K and C are finite. Say $|Q|=n$, $K=\{1,\ldots,k\}$ and $A=K\cup C=\{1,\ldots,m\}$ for some integers n,k,m with $n\geq 0$ and $m\geq k\geq 1$, and say B is a set of colours. (Readers can generalise the visual image given below as much as they like. The situation is conceptually the same for general K,C,B and Q.) Consider the $|Q|$-dimensional Euclidean space $\mathbb{R}^Q$, with coördinate axes labelled by the elements of Q. Then A^Q is the subset of $\mathbb{R}^Q$ consisting of those points whose coordinates are all in A, and these points form an $m\times m\times\ldots\times m$ (n times) *grid* in this space. (Note that these are called points in the usual geometric sense, and should not be confused with the elements of Q, which can also be called points. The elements of Q label the n axes, whereas there are m^n points in the grid.) To each point $x\in A^Q$, in this grid, the colour $f(x)\in B$ is assigned, creating a *coloured grid* sitting in $\mathbb{R}^Q$.

To make $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ an isomorphism class it is necessary to define structure isomorphism, and this is done "naturally" (see section 3). For all ground sets $Q,P\in\mathscr{Q}$, any function $f:A^Q\to B$ (that is, $f\in B^{(A^Q)}$) and any bijection $\omega:Q\to P$, the function $\omega(f):A^P\to B$ (that is, $\omega(f)\in B^{(A^P)}$), isomorphic to f via ω, is defined "naturally" as follows. In the above visualisation, the q–axis is simply renamed as the $\omega(q)$–axis for each $q\in Q$, so that the "coloured grid", while itself unchanged, now sits in $\mathbb{R}^P$. More formally (but equivalently) $\omega:A^Q\to A^P$ and $\omega:B^{(A^Q)}\to B^{(A^P)}$ are defined as follows. For $x\in A^Q$ (that is, $x:Q\to A$) let $\omega(x)\in A^P$ (that is, $\omega(x):P\to A$) satisfy $[\omega(x)](\omega(q))=x(q)$ for all $q\in Q$ (so that $\omega(x)$ is actually $x\circ\omega^{-1}$). (Recall that $\omega(\mathscr{R})$ was defined in this way in section 3.) For $f\in B^{(A^Q)}$ (that is, $f:A^Q\to B$) let $\omega(f)\in B^{(A^P)}$ (that is, $\omega(f):A^P\to B$) satisfy $[\omega(f)](\omega(x))=f(x)$ for all $x\in A^Q$. Since isomorphism is defined naturally, conditions (M1) and (M2) automatically hold, (although it is routine to show this formally).

To make $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ a minor clas (and a point–removal class when all ground sets are finite), it is necessary to define point removal. The visualisation of point removal is much more illuminating than the formal definition (as was the case with structure isomorphism) and so it is given first. Recall the coloured grid in $\mathbb{R}^Q$, associated with $f:A^Q\to B$. For any point $q\in Q$ and any manner $\ell\in K$, consider the $m\times m\times\ldots\times m$ (n–1 times, where $n=|Q|$) coloured subgrid consisting of those gridpoints with q^{th} coordinate ℓ. (This is the "cross section" of the coloured grid taken by the hyperplane of $\mathbb{R}^Q$, perpendicular to the q–axis and intersecting this axis at coordinate ℓ.) This coloured subgrid can be "projected" into $\mathbb{R}^{Q-\{q\}}$ by sending each gridpoint $(x_t\,|\,t\in Q)\in A^Q$ with $x_q=\ell$, to the gridpoint $(x_t\,|\,t\in Q-\{q\})\in A^{Q-\{q\}}$, discarding the q^{th} coordinate. With each such gridpoint retaining its colour, this yields a coloured grid sitting in $\mathbb{R}^{Q-\{q\}}$, which depicts the structure on ground set

$Q-\{q\}$ (a function from $A^{Q-\{q\}}$ to B) obtained from f by removing point q in manner ℓ. (Multiple point removals are defined soon.) It is clear that this definition of point removal respects isomorphism so that (M3') holds. Also for any distinct points $q,r \in Q$ and any manners $\ell,h \in K$ the coloured grid sitting in $\mathbb{R}^{Q-\{q,r\}}$ obtained by removing point q in manner ℓ and point r in manner h (in either order) is obtained from the coloured grid for f sitting in $\mathbb{R}^Q$, by taking the $m \times m \times \ldots \times m$ (n–2 times) coloured subgrid consisting of those gridpoints with q^{th} coordinate ℓ and r^{th} coordinate h, and projecting it into $\mathbb{R}^{Q-\{q,r\}}$ (while preserving grid point colour) by discarding the q^{th} and r^{th} coordinates. (Note that this (n–2)–dimensional subgrid is the intersection of the two (n–1)–dimensional subgrids associated with the two removals of a single point.) This shows that (M4') holds.

More generally, for any prescription $\mathfrak{K} \in \overline{K}^Q$, consider the $m \times m \times \ldots \times m$ ($|G(\mathfrak{K})|$ times) coloured subgrid consisting of those gridpoints with q^{th} coordinate $\mathfrak{K}(q)$ for all points $q \in Q-G(\mathfrak{K})$. (This is the intersection of all the subgrids associated with removing a single point $q \in Q-G(\mathfrak{K})$ in manner $\mathfrak{K}(q)$.) This coloured subgrid can be projected into $\mathbb{R}^{G(\mathfrak{K})}$ by sending each gridpoint $(x_t \mid t \in Q) \in A^Q$ with $x_q = \mathfrak{K}(q)$ for all $q \in Q-G(\mathfrak{K})$, to the gridpoint $(x_t \mid t \in G(\mathfrak{K})) \in A^{G(\mathfrak{K})}$, discarding the q^{th} coordinate for all $q \in Q-G(\mathfrak{K})$. With each such gridpoint retaining its colour, this yields a coloured grid sitting in $\mathbb{R}^{G(\mathfrak{K})}$, which depicts the structure obtained from f by removing point q in manner $\mathfrak{K}(q)$ for all $q \in Q-G(\mathfrak{K})$, (as prescribed by $\mathfrak{K}$). Again, point removal respects isomorphism so that (M3) holds, and it should be clear that (M4) and (M5) also hold (see the interpretations below the statements of these conditions), making $\mathcal{F}^{\mathcal{Q}}(K,C,B)$ a minor class. (This is shown formally below.)

Here is the formal definition of point removal. For any structure f in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$, on ground set Q, that is $f: A^Q \to B$, and any prescription $\mathfrak{K} \in \overline{K}^Q$, the structure $f[\mathfrak{K}]$ on ground set $G(\mathfrak{K})$ is the following function from $A^{G(\mathfrak{K})}$ to B. For every $x=(x(q)\,|\,q\in G(\mathfrak{K}))\in A^{G(\mathfrak{K})}$ it holds that $(f[\mathfrak{K}])(x)=f(\mathfrak{K}\Delta x)$. (See section 3 for the definition of Δ.) With prescription $\mathfrak{N}$ as in condition (M4), it follows that $\mathfrak{N}\Delta x=x$ for all $x\in A^Q$ and hence $(f[\mathfrak{N}])x=f(\mathfrak{N}\Delta x)=f(x)$ so that $f[\mathfrak{N}]=f$ and (M4) holds. With prescriptions $\mathfrak{J},\mathfrak{L}$ as in condition (M5), it is clear that $(\mathfrak{J}\Delta\mathfrak{L})\Delta x=\mathfrak{J}\Delta(\mathfrak{L}\Delta x)$ for all $x\in A^{G(\mathfrak{L})}$. It follows that $(f[\mathfrak{J}\Delta\mathfrak{L}])(x)=f((\mathfrak{J}\Delta\mathfrak{L})\Delta x)=f(\mathfrak{J}\Delta(\mathfrak{L}\Delta x))=(f[\mathfrak{J}])(\mathfrak{L}\Delta x)=((f[\mathfrak{J}])[\mathfrak{L}])(x)$ so that $f[\mathfrak{J}\Delta\mathfrak{L}]=(f[\mathfrak{J}])[\mathfrak{L}]$ and (M5) holds. Clearly conditions (M1),(M2) and (M3) hold (it is routine to show this) so that $\mathcal{F}^{\mathcal{Q}}(K,C,B)$ is indeed a minor class.

The "coloured grid" visualisation appears in later sections, including some proofs, to provide insight into a situation, but it is never a formal part of a proof.

SECTION 5: MINOR CLASSES AS ALGEBRAS

Introduction

In this section, it is shown how minor classes can be treated as algebras, making much of the theory of universal algebra available to them. Some basic universal algebra definitions are given, along with minor class examples showing their relevance.

For convenience it will be required, until further notice, that a minor class $(\mathscr{S},\mathscr{Q},\mathscr{I},K,\mathscr{P})$ must have $\mathscr{S},\mathscr{Q}$,K and hence $\mathscr{I}$ and $\mathscr{P}$ being sets. I suggest that this involves no loss of generality. (While this appears less general, it is equivalent (by changing the word "set" to "class") to the case where $\mathscr{Q}$ is a class of <u>classes</u>, which appears more general. See the discussion in section 2.) Readers can draw their own conclusions.

Universal algebra originated with Birkhoff's 1935 paper [1], which gives a general definition of what an "algebra" is, (covering many well known mathematical entities of the time), and develops a large body of theory. Higgins' 1963 paper [6] extends this definition (admitting other well known mathematical entities as algebras), and shows that the theory can be extended similarly. Higgins' "algebras" are called *many-sorted* or *heterogeneous* while those of Birkhoff are called *1-sorted* or *homogeneous.* By Higgins' definition, a (many-sorted) algebra A consists of a $\mathscr{Q}$-partition $A=\langle A_s | s\in\mathscr{Q}\rangle$ (for some set $\mathscr{Q}$) together with a set of functions, each of the form $A_{s_1}\times\cdots\times A_{s_n}\to A_s$ for some integer $n\geq 0$ and sorts $s_1,\cdots,s_n,s\in\mathscr{Q}$. (Birkhof's definition has $|\mathscr{Q}|=1$.) When n=1 the function is *unary*, and when all the functions are unary, the algebra is *unary.*

As shown later, minor classes are algebras (in fact, many-sorted unary algebras). This means that there is a large body of established theory that can be applied to them. Unfortunately, most universal algebra is still done for the 1-sorted case, but fortunately, most results generalise to the many sorted case.

When universal algebraists specialise, minor classes are often among the algebras to miss out, since they are very different from algebras that are commonly studied. Nonetheless, unary algebras (including minor classes) are easier to deal with than algebras in general, making it easier to develop specialised theory for them.

Given below are some standard universal algebra definitions, with examples putting minor classes into context. The definitions only cover the unary case, since that is all that is needed here, and the notation is simpler. Full definitions appear in [3], while [5] contains more theory (but only for 1–sorted algebras).

Definitions

A *unary signature* is a pair $\Sigma=(\mathcal{Q},\mathcal{O})$ where $\mathcal{Q}$ is a set whose elements are called *sorts* and $\mathcal{O}=<\mathcal{O}_{s,t}|(s,t)\in\mathcal{Q}\times\mathcal{Q}>$ is a $\mathcal{Q}\times\mathcal{Q}$–partition of a set whose elements are called *operator symbols.* (It is permissible that some $\mathcal{O}_{s,t}$ may be empty.)

A (*unary*) *algebra* **A** *of signature* Σ, or a Σ*–algebra*, consists of a $\mathcal{Q}$–partition of a set $A=<A_s|s\in\mathcal{Q}>$, the *universe* of A, together with a set of functions; one function $f_{\mathbf{A}}:A_s\to A_t$ for all sorts $s,t\in\mathcal{Q}$ and every operator symbol $f\in\mathcal{O}_{s,t}$. (It is permissible that some A_s, or some $\mathcal{O}_{s,t}$, may be empty.) Except for minor classes (which are denoted by script letters) any algebra is denoted by the boldface version of the letter used for its universe. (For example, the universe of **B** is always $B=<B_s|s\in\mathcal{Q}>$.)

A minor class $(\mathcal{S},\mathcal{Q},\mathcal{J},K,\mathcal{P})$ can be formulated as a unary algebra where the sorts are precisely the ground sets; that is, as a Σ–algebra where $\Sigma=(\mathcal{Q},\mathcal{O})$ with $\mathcal{O}$ to be determined. For all sorts (ground sets) $Q,P\in\mathcal{Q}$, let $\mathcal{O}_{Q,P}$ consist of operator symbols i^{ω} for each bijection $\omega:Q\to P$ and $p^{\mathfrak{K}}$ for each prescription $\mathfrak{K}\in\overline{K}^{Q}$ with $G(\mathfrak{K})=P$. (Observe that $\mathcal{O}_{Q,P}$ is non–empty only if $|Q|=|P|$ or $P\subseteq Q$.) So the unary function $i^{\omega}_{\mathcal{S}}:\mathcal{S}_Q\to\mathcal{S}_P$ sends each $S\in\mathcal{S}_Q$ to $\omega(S)\in\mathcal{S}_P$ and

$p^{\mathfrak{K}}_{\mathscr{S}}:\mathscr{S}_Q \to \mathscr{S}_P$ sends each $S \in \mathscr{S}_Q$ to $S[\mathfrak{K}] \in \mathscr{S}_P$. It is convenient to call operator symbols of the form i^{ω}, and also the corresponding unary functions, (*structure*) *isomorphisms*, and to call operator symbols of the form $p^{\mathfrak{K}}$, and also the corresponding unary functions, *point removals*. (Note that a point removal can correspond to removing all, some, one or none of the points in the ground set of a structure.) So as a unary algebra, the minor class $(\mathscr{S}, \mathscr{Q}, \mathscr{I}, K, \mathscr{P})$ has universe $\mathscr{S} = \langle \mathscr{S}_Q \mid Q \in \mathscr{Q} \rangle$ with functions in $\mathscr{I}$ and $\mathscr{P}$. The signature $(\mathscr{Q}, \mathscr{O})$ is determined by $\mathscr{Q}$ and K. Observe however, that unary algebras with this signature need not be minor classes, since they need not satisfy conditions (M1) to (M5). Fortunately these conditions are of a form which universal algebra handles best, namely equations. These are defined below.

Equations and Varieties

Let $\Sigma = (\mathscr{Q}, \mathscr{O})$ be a unary signature and consider a $\mathscr{Q}$-partition of a set $X = \langle X_s \mid s \in \mathscr{Q} \rangle$. The elements of X_s shall be called *variables of sort* s, for each sort $s \in \mathscr{Q}$. A *Σ-term of sort* s_0 *in variables* X is a (possibly empty) string of operator symbols followed by a variable, $f^1 f^2 \ldots f^n x$ for some integer $n \geq 0$ such that operator symbols $f^i \in \mathscr{O}_{s_{i-1}, s_i}$ for $i = 1, \cdots, n$ and variable $x \in X_{s_n}$ for some sorts $s_0, s_1, \cdots, s_n \in \mathscr{Q}$. In particular if $n=0$, this says that $x \in X_{s_0}$ is a term of sort s_0. Let $T^{\Sigma}_s(X)$ be the set of Σ-terms of sort s in variables X and let $T^{\Sigma}(X) = \langle T^{\Sigma}_s(X) \mid s \in \mathscr{Q} \rangle$. For all $Y \subseteq X$, define $T^{\Sigma}_s(Y)$ and $T^{\Sigma}(Y)$ similarly.

For any $Y \subseteq X$ and any Σ-algebra **A**, an *assignment* is a $\mathscr{Q}$-respecting function $\alpha: Y \to A$. The *extension* of α is the $\mathscr{Q}$-respecting function $\overline{\alpha}: T^{\Sigma}(Y) \to A$ which sends each term $f^1 f^2 \cdots f^n x \in T^{\Sigma}_{s_0}(Y)$ to $f^1_{\mathbf{A}}(f^2_{\mathbf{A}}(\ldots(f^n_{\mathbf{A}}(\alpha(x))\cdots)) \in A_{s_0}$. (The definition of term ensures that this last expression makes sense, that is, the unary functions are only applied to elements on which they act.)

A *Σ-equation of sort* s is a triple (Y, ℓ, r) where $Y \subseteq X$ and $\ell, r \in T^{\Sigma}_s(Y)$. The equation (Y, ℓ, r) is *valid in a* Σ-algebra **A** if for every assignment $\alpha: Y \to A$ it holds that $\overline{\alpha}(\ell)=\overline{\alpha}(r)$. Observe that if there were some sort $s\in\mathcal{S}$ with $Y_s\neq\emptyset$ but $A_s=\emptyset$, then the equation would automatically be valid, since there would be no assignment α:Y→A. It follows that adding extra variables to Y can change the validity of the equation in some algebras.

Define var(ℓ,r) to be the set of variables which appear in term ℓ or term r (there are one or two of them). The equation (Y,ℓ,r) is a *proper equation* if Y=var(ℓ,r) and in this case is abbreviated to the pair (ℓ,r), usually written ℓ=r. Almost all equations in the literature are proper. The proper equation ℓ=r is a *regular equation* if var(ℓ,r) has only one element, that is, the same variable appears on both sides.

An *equational specification* (respectively *proper* equational specification, *regular* equational specification) is a pair $\Lambda=(\Sigma,\Theta)$ where Θ is a set of Σ–equations (respectively proper Σ–equations, regular Σ–equations). A Σ–algebra **A** is an *algebra of the specification* Λ, or a Λ*–algebra*, if every equation in Θ is valid in **A**. (Note that Σ can be treated as a regular equational specification in which Θ is empty.) A class of algebras is a *variety* (respectively *proper* variety, *regular* variety) if it is the class of all Λ–algebras for some equational specification Λ (respectively proper equational specification Λ, regular equational specifications Λ). The adjective *unary* can also precede the word variety, if it is to be emphasised that it is a variety of unary algebras. It should be noted that all varieties of 1–sorted algebras are proper varieties, but this does not hold for many–sorted algebras.

For any set of Σ–equations Θ, let alg(Θ) be the class of all (Σ,Θ)–algebras. So all (proper, regular) varieties of Σ–algebras are of the form alg(Θ) for some set Θ of (proper, regular) Σ–equations. For any class $\mathscr{C}$ of Σ–algebras let eq($\mathscr{C}$) (respectively peq($\mathscr{C}$), req($\mathscr{C}$) be the set of Σ–equations (respectively proper Σ–equations, regular Σ–equations) which are valid in every algebra in $\mathscr{C}$. The smallest variety (respectively proper variety, regular variety) containing $\mathscr{C}$

is alg(eq($\mathscr{C}$)) (respectively alg(peq($\mathscr{C}$)), alg(req($\mathscr{C}$))). If $\mathscr{C}$ is a proper variety then the equations in eq($\mathscr{C}$) are exactly those of the form (Y,ℓ,r) for proper equations $(\ell,r)\in$peq($\mathscr{C}$). If $\mathscr{C}$ is a regular unary variety then all proper equations in peq($\mathscr{C}$) are regular, that is peq($\mathscr{C}$)=req($\mathscr{C}$). Incidentally, eq(alg(eq($\mathscr{C}$)))=eq($\mathscr{C}$) for every $\mathscr{C}$ and alg(eq(alg(Θ)))=alg(Θ) for every Θ.

A set of Σ-equations of the form eq($\mathscr{C}$) is *deductively closed.* There are rules [3] by which any equation in eq(alg(Θ)), the *deductive closure* of Θ, can be derived from equations in Θ. (These correspond to the way equations are manipulated in practice.) For all sorts $s\in\mathscr{Q}$, let $\mathrm{eq}_s^Y(\mathscr{C})$ be the set of all pairs (ℓ,r) where the equation (Y,ℓ,r), of sorts, is in eq($\mathscr{C}$) (that is, is valid in every algebra in $\mathscr{C}$) and let $\mathrm{eq}^Y(\mathscr{C})$ be $<\mathrm{eq}_s^Y(\mathscr{C})|s\in\mathscr{Q}>$. Then $\mathrm{eq}^Y(\mathscr{C})$ is a $\mathscr{Q}$–respecting equivalence relation on $T^\Sigma(Y)$. So two terms $\ell,r\in T_s^\Sigma(Y)$, of sort s, are equivalent by $\mathrm{eq}^Y(\mathscr{C})$ exactly when the equation (Y,ℓ,r) is valid in every algebra in $\mathscr{C}$. In particular when $\mathscr{C}$=alg(Θ), the term r can be obtained from the term ℓ (and visa–versa) by a sequence of formal manipulations [3] involving the equations in Θ. (Again these manipulations correspond to what is done in practice.)

The truth of conditions (M1)–(M5), imposed on minor classes, is equivalent to the validity of all of a certain set of regular unary equations of the appropriate signature (which depends on $\mathscr{Q}$ and K). For example, by simple conversion of notation, condition (M5) is equivalent to the validity of all equations $p^{\mathfrak{J}\Delta\mathfrak{L}}x=p^{\mathfrak{L}}p^{\mathfrak{J}}x$, for every sort $Q\in\mathscr{Q}$, some variable x of sort Q, and every $\mathfrak{J}$ and $\mathfrak{L}$ as given in condition (M5). Let EM5 be the set of these equations. Similarly, with the superscripts below taking all possible values as given in the corresponding condition, and x being a variable of the appropriate sort in each case (there is a different variable for each sort), let EM1 consist of all equations $i^1x=x$, let EM2 consist of all equations $i^\tau i^\omega x=i^{\tau\circ\omega}x$, let EM3 consist of all equations $i^{\omega|G(\mathfrak{K})}p^{\mathfrak{K}}x=p^{\omega(\mathfrak{K})}i^{\omega}x$ and let EM4 consist of all equations $p^{\mathfrak{N}}x=x$. Let EM be the union of EM1 to EM5. Note that the universal quantifier "for every structure $s\in\mathcal{Q}_Q$" is not mentioned when listing these equations, since it is incorporated in the definition of validity of these equations. (See the definition of "valid" given earlier.)

It is convenient to call a minor class, with set of ground sets $\mathcal{Q}$ and set of manners K, a $(\mathcal{Q},K)$ minor class. Then the class of all $(\mathcal{Q},K)$ minor classes, is the class of all algebras of the appropriate signature, in which all the (regular unary) equations in EM are valid, so that it is a regular unary variety. There is a separate variety for each distinct pair $(\mathcal{Q},K)$. Similarly there is a (regular unary) variety of structure classes, and one of isomorphism classes, for each $\mathcal{Q}$, and there is a (regular unary) variety of pre point–removal classes, and one of point–removal classes, for each $(\mathcal{Q},K)$.

Any minor class term will consist of a variable preceded by a (possibly empty) string of isomorphisms and point removals. The term manipulations, mentioned above, allow any such term to be put into a standard form. Firstly, the equations in EM3 allow all the isomorphisms to be taken to the left of the point removals. Then the equations in EM1 and EM2 allow any (possibly empty) string of isomorphisms to be reduced to one isomorphism, and the equations in EM4 and EM5 do the same for point removals. So any minor class term of sort $P\in\mathcal{Q}$, with variable x of sort $Q\in\mathcal{Q}$, is equivalent (in the presence of the equations in EM) to the *canonical term* $i^{\omega}p^{\mathfrak{K}}x$ for some prescription $\mathfrak{K}\in\overline{K}^{Q}$ and some bijection $\omega{:}G(\mathfrak{K})\rightarrow P$. (Note that $\mathfrak{K}$ and ω are uniquely determined by the original term.) So applying any sequence of minor class functions (isomorphisms and point removals) to a structure, is equivalent to applying one point removal followed by one isomorphism.

Subalgebras, Homomorphisms and Direct Products

In [1], Birkhoff gives two equivalent characterisations of when a class of Σ-algebras (for some signature Σ) is a variety. Higgins [6] extends this to many-sorted algebras. One characterisation is the definition given earlier, while the other is that the class of algebras is closed under certain constructions (by which some algebras are constructed from others). These three constructions, namely, subalgebras, homomorphic images, and direct products, are given below, together with related theory. The above equivalence ensures that when these constructions are applied to $(\mathcal{Q},\mathrm{K})$ minor classes (for some $\mathcal{Q}$ and K) they yield $(\mathcal{Q},\mathrm{K})$ minor classes (since these form a variety).

Subalgebras

Let $\Sigma=(\mathcal{Q},\mathcal{O})$ be a unary signature and let **A** and **B** be Σ-algebras. Then **B** is a *subalgebra* of **A**, and **A** is an *extension* of **B** if $\mathrm{B}\subseteq\mathrm{A}$ and $f_{\mathbf{A}}(a)=f_{\mathbf{B}}(a)$ for all sorts $s,t\in\mathcal{Q}$, every operator symbol $f\in\mathcal{O}_{s,t}$ and every *element* $a\in\mathrm{B}_s$. In this case we write $\mathbf{B}\leq\mathbf{A}$ and $\mathrm{B}\leq\mathrm{A}$ and say that B is a *subuniverse* of A. Equivalently, B is a subuniverse of A if $\mathrm{B}\subseteq\mathrm{A}$ and $f_{\mathbf{A}}(a)\in\mathrm{B}_t$ for all sorts $s,t\in\mathcal{Q}$, every operator symbol $f\in\mathcal{O}_{s,t}$ and every element $a\in\mathrm{B}_s$, and **B** is the unique *subalgebra* (of A) with universe B, where each function $f_{\mathbf{B}}:\mathrm{B}_s\to\mathrm{B}_t$ is defined by setting $f_{\mathbf{B}}(a)=f_{\mathbf{A}}(a)$ for all such s,t,f and a.

For example, if $\mathscr{S}$ is a $(\mathcal{Q},\mathrm{K})$ minor class (for some $\mathcal{Q}$ and K) and $\mathscr{T}$ is a set of structures, then $\mathscr{T}$ is a subuniverse of $\mathscr{S}$ provided that $\mathscr{T}\subseteq\mathscr{S}$ and for all ground sets $\mathrm{Q},\mathrm{P}\in\mathcal{Q}$, every prescription $\mathfrak{K}\in\overline{\mathrm{K}}^{\mathrm{Q}}$, every bijection $\omega:\mathrm{Q}\to\mathrm{P}$, and every structure $\mathrm{S}\in\mathscr{T}_{\mathrm{Q}}$, it holds that the structures $\mathrm{S}[\mathfrak{K}]$ and $\omega(\mathrm{S})$ (as defined in $\mathscr{S}$) are also in $\mathscr{T}$. That is, $\mathscr{T}$ is closed under isomorphism and point removal, or equivalently, closed under isominors. Alternatively, a $(\mathcal{Q},\mathrm{K})$ minor class $\mathscr{T}$ is a subalgebra of $\mathscr{S}$, provided $\mathscr{T}\subseteq\mathscr{S}$ and the definition of the functions in $\mathscr{T}$ (point removal and isomorphism) on any structure in $\mathscr{T}$, coincides with the definition of these functions in $\mathscr{S}$, on structures in $\mathscr{T}$. In this case, $\mathscr{T}$ is a *sub minor class* of $\mathscr{S}$.

For any Σ–algebra **A**, the partial order ≤ on subuniverses of A, induces a complete lattice, so that the meet and join of any set of subuniverses is defined. (See section 2.) The same is true for subalgebras, since there is an obvious one–to–one correspondence between subalgebras and subuniverses. The meet of subuniverses is the intersection of subuniverses. For any C⊆A, there exists a unique subalgebra of **A**, denoted **A**incC, such that the corresponding subuniverse, denoted AincC, minimally contains C. That is AincC is the intersection of all subuniverses containing C. If **A**incC is **A**, then **A** *is generated by* C (or C *generates* **A**).

What appears in the previous paragraph holds for all algebras. However, a property which rarely holds for algebras in general, but which holds for all unary algebras, is that the join of subuniverses is the union of subuniverses. This is the major reason why unary algebras are simpler to study than algebras in general. For any Σ–algebra **A** (where Σ is a unary signature) and any C⊆A, there exists a unique subalgebra of **A**, denoted **A**excC, such that the corresponding subuniverse, denoted AexcC, is maximally disjoint from C. That is, AexcC is the union of all subuniverses disjoint from C.

Define a relation ≤ on elements of A (this cannot be confused with the partial order ≤ on subuniverses of A) where for all elements a,b∈A it holds that a≤b exactly when a∈Ainc{b} or equivalently b∉Aexc{a}. That is, every subuniverse containing element b, also contains element a. An equivalent definition, independent of the previous two paragraphs, is given inductively by saying that a≤a for every element a∈A and for all elements a,b∈A, if a≤b then $f_{\mathbf{A}}(a) \leq b$ for any $f_{\mathbf{A}}$ which acts on element a, (that is, if $a \in A_s$ for some sort s then the operator symbol $f \in \mathcal{O}_{s,t}$ for some sort t). So a≤b exactly when element a can be obtained from element b by repeated application of the unary functions of A, and Ainc{b} consists of all such elements a.

This relation $\leq$, on elements of A, is a *quasi order* (that is, it is reflexive and transitive). Let $a\simeq b$ be defined to mean $a\leq b$ and $b\leq a$, so that $\simeq$ is clearly an equivalence relation. Let $a<b$ be defined to mean $a\leq b$ and $b\nleq a$ (or equivalently, $a\leq b$ and $a\not\simeq b$). For any $C\subseteq A$, observe that AincC consists of all elements $a\in A$ such that $a\leq b$ for some $b\in C$ and that AexcC consists of all elements $a\in A$ such that $b\nleq a$ for every element $b\in C$.

If $a\leq b$ then element a is an *isominor* of element b. For minor classes, this is equivalent to the earlier definition; since as shown before, any structure obtained from another structure by a sequence of minor class functions, can be obtained by performing a point removal followed by an isomorphism. For any two structures S and T in a minor class with only finite ground sets, $S\simeq T$ if and only if $S\cong T$. (Since $S\simeq T$ means that both S and T are isomorphic to a minor of the other, so that $|G(S)|\leq|G(T)|$ and $|G(T)|\leq|G(S)|$ and hence $|G(S)|=|G(T)|$. Since ground sets are finite, neither S nor T is isomorphic to a proper minor of the other, so that $S\cong T$. Also $S\cong T$ clearly implies $S\simeq T$.) This need not hold when there are infinite ground sets. For example, consider the minor class $\mathscr{W}^{\mathscr{Q}}$ (see section 3) such that there exists a countably infinite ground set $Q\in\mathscr{Q}$. Define structures $W_1,W_2\in\mathscr{W}^{\mathscr{Q}}$, on ground set Q, where $W_1=\{P|P\subseteq Q$ and $|P|$ is odd$\}$ and $W_2=\{P|P\subseteq Q$ and $|P|$ is even$\}$. Then contracting any one point from one of them gives a proper minor, isomorphic to the other, while W_1 is not isomorphic to W_2. That is, $W_1\simeq W_2$ while $W_1\ncong W_2$.

Given a Σ-algebra A and a subuniverse B of A, we often wish to find $C\subseteq A$ such that B=AexcC, and to express C economically. An *infinite* (*strictly*) *descending chain* in A is an infinite sequence of elements $a_1,a_2,a_3,\cdots$ of A such that $\cdots<a_3<a_2<a_1$. If such a sequence exists then **A** *has an infinite descending chain*, otherwise **A** *has no infinite descending chain.* Suppose **A** has

the above infinite descending chain, and that B=Aexc$\{a_1,a_2,a_3,\cdots\}$. Then for any C such that B=AexcC, it must be that C has redundant elements which can be removed from C without changing the subuniverse. (That is, there exists D$\subseteq$C with D$\neq$C and B=AexcD.) This is because any element of C is strictly above ($>$) some a_i while some other element must be below ($\leq$) a_i making the former element redundant.

An element b in a quasi ordered set is *minimal* if for any element a with a$\leq$b it holds that b$\leq$a, so that a$\simeq$b. Now if A has no infinite descending chain (as is the case for minor classes with only finite ground sets) then every element of A–B is bounded below by a minimal element of A–B. If element b is minimal in A–B then so is every element a$\in$A–B such that a$\simeq$b. Let b$\simeq$ be the set $\{a \mid a\in$B–A and a$\simeq$b$\}$ of such elements. If B=AexcC then C must contain at least one element of b$\simeq$ (any one will do) but it need not contain more than one (since the exclusion of any element a$\in$(b$\simeq$) from A, removes every element c$\geq$a and in particular every element c$\in$(b$\simeq$)). So if C consists of exactly one element of b$\simeq$ for each element b, minimal in A–B, then C has no redundant elements.

A minor class $\mathscr{S}$, with only finite ground sets, has no infinite descending chain $\cdots<S_3<S_2<S_2<S_1$, since this would require an infinite descending chain $\cdots<|G(S_3)|<|G(S_2)|<|G(S_1)|$ of positive integers – an impossibility. So for any sub minor class $\mathscr{T}$ of $\mathscr{S}$, any structure in $\mathscr{S}-\mathscr{T}$ is bounded below by a minimal structure in $\mathscr{S}-\mathscr{T}$. These minimal structures are the *excluded isominors of* $\mathscr{T}$ *in* $\mathscr{S}$. Let C$\subseteq\mathscr{S}-\mathscr{T}$ contain exactly one of each of these (any one will do) up to isomorphism, or equivalently, as shown earlier, up to $\simeq$-equivalence. Then $\mathscr{T}=\mathscr{S}$excC and C has no redundant elements. The elements of C are the *excluded minors of* $\mathscr{T}$ *in* $\mathscr{S}$. These excluded minors are determined up to isomorphism. Often an excluded minor is described, not as a particular structure, but only in a way which determines it up to isomorphism, which is all that is needed anyway.

Recall the minor class $\mathscr{W}$, defined in section 3, which has only finite ground sets. Let $\mathscr{W}^U$ consist of those structures $(W,Q)\in\mathscr{W}$ (recall that $W\subseteq 2^Q$) which have the property that if $P\in W$ and $P\subseteq R\subseteq Q$, then $R\in W$. Equivalently, if $P\in W$ and $q\in Q-P$, then $P\cup\{q\}\in W$. It follows that if $(W,Q)\in(\mathscr{W}-\mathscr{W}^U)$ then there exists $P\in W$ and $q\in Q-P$ such that $P\cup\{q\}\notin W$. Contracting every point in P from W, and deleting all other points except q leaves $(\{\emptyset\},\{q\})$ which is also in $\mathscr{W}-\mathscr{W}^U$. This shows that $\mathscr{W}^U=\mathscr{W}\,\mathrm{exc}\{(\{\emptyset\},\{q\})\}$, where $\{q\}$ is an arbitrary one element ground set. So $\mathscr{W}^U$ is a sub minor class of $\mathscr{W}$ with a single excluded minor (in $\mathscr{W}$).

Recall the minor class $\mathscr{F}^{\mathscr{Q}}(K,C,B)$, defined in section 4. For any $D\subseteq B$ the minor class $\mathscr{F}^{\mathscr{Q}}(K,C,D)$ is a sub minor class of $\mathscr{F}^{\mathscr{Q}}(K,C,B)$. However, we cannot talk of excluded minors when there exist infinite ground sets.

Homomorphisms

Let $\Sigma=(\mathscr{Q},\mathscr{O})$ be a unary signature and let **A** and **B** be Σ-algebras. A *Σ-homomorphism from* **A** *to* **B** is a $\mathscr{Q}$-respecting function $\alpha:A\to B$, with $\mathscr{Q}$-partition $<\alpha_s:A_s\to B_s\,|\,s\in\mathscr{Q}>$, such that *$\alpha$ respects* f, that is, $\alpha_t(f_{\mathbf{A}}(a))=f_{\mathbf{B}}(\alpha_s(a))$ for all sorts $s,t\in\mathscr{Q}$, every operator symbol $f\in\mathscr{Q}_{s,t}$ and every element $a\in A_s$. Observe that this condition ensures that if $C\subseteq A$ and C generates **A**, then specifying the effect of α on the elements of C, determines its effect on all of A.

If $\mathscr{S}$ and $\mathscr{T}$ are $(\mathscr{Q},K)$ minor classes (for some set of ground sets $\mathscr{Q}$ and set of manners $\mathscr{Q}$) then $\alpha:\mathscr{S}\to\mathscr{T}$ is a minor class homomorphism provided that the following three conditions hold. Firstly, *α respects ground sets* (sorts), that is, $G(\alpha(S))=G(S)$ for every structure $S\in\mathscr{S}$. Secondly, *α respects point removals*, that is $\alpha(S[\mathfrak{K}])=(\alpha(S))[\mathfrak{K}]$ for every structure $S\in\mathscr{S}$ and every prescription $\mathfrak{K}\in\overline{K}^{G(S)}$. (Note that the point removal on the left hand side acts on structures in $\mathscr{S}$ while the point removal on the right hand side acts on structures in $\mathscr{T}$. The same goes for the structure isomorphism in the next condition.) Thirdly, α *respects* (*structure*) *isomorphisms*, that is, $\alpha(\omega(S))=\omega(\alpha(S))$ for every structure $S\in\mathscr{S}$, every ground set $P\in\mathscr{Q}$ and every bijection ω:$G(S)\to P$.

Let **A**,**B**,**C** be Σ-algebras and let α:**A**$\to$**B** and β:**B**$\to$**C** be Σ-homomorphisms. The *composition* of homomorphisms α and β is $(\beta\circ\alpha)$:**A**$\to$**C**, that is, $<(\beta_s\circ\alpha_s)\,|\,s\in\mathscr{Q}>$. Clearly $\beta\circ\alpha$ is a Σ-homomorphism from **A** to **C**. The *homomorphic image* α(**A**) *of* **A** *under* α is the subalgebra of **B** with universe $\alpha(A)=<\alpha_s(A_s)\,|\,s\in\mathscr{Q}>$ where $\alpha_s(A_s)=\{\alpha_s(a)\,|\,a\in A_s\}$. The homomorphism α:**A**$\to$**B** is an *isomorphism* if α is a bijection (that is, α_s is a bijection for each sort $s\in\mathscr{Q}$). In this case **A** *is isomorphic to* **B**, denoted **A**$\cong$**B**. An *embedding* is an injective homomorphism α:**A**$\to$**B** (that is, α_s is injective for each sort $s\in\mathscr{Q}$). In this case, **A** *can be embedded in* **B**. (This is equivalent to saying that **A** is isomorphic to a subalgebra of **B**, namely α(**A**).)

A *congruence* on **A** is a $\mathscr{Q}$-respecting equivalence relation $q\subseteq A\times A$, with $\mathscr{Q}$-partition $<q_s\subseteq A_s\times A_s\,|\,s\in\mathscr{Q}>$, such that aq_sb implies $(f_{\mathbf{A}}(a))q_t(f_{\mathbf{A}}(b))$ for all sorts $s,t\in\mathscr{Q}$, every operator symbol $f\in\mathscr{O}_{s,t}$ and all elements $a,b\in A_s$. For every sort $s\in\mathscr{Q}$ and every element $a\in A_s$, let aq be $\{b\,|\,b\in A_s \text{ and } aqb\}$, let A_s/q be $\{aq\,|\,a\in A_s\}$, and let $A/q=<A_s/q\,|\,s\in\mathscr{Q}>$. The algebra **A**/q, with universe A/q, is defined by setting $f_{\mathbf{A}/q}(aq)=(f_{\mathbf{A}}(a))q$ for all sorts $s,t\in\mathscr{Q}$, every operator symbols $f\in\mathscr{O}_{s,t}$, and every element $a\in A_s$. The conditions on q ensure that this is well defined. The function which sends each $a\in A$ to $(aq)\in A/q$ is a homomorphism, and the homomorphic image of **A** is **A**/q.

The *kernel* of a homomorphism α:**A**$\to$**B** is the $\mathscr{Q}$-respecting equivalence relation ker(α) on A, for which $(a,b)\in\ker(\alpha)$ exactly when $\alpha(a)=\alpha(b)$, for all elements $a,b\in A$. A relation on the universe A, of **A**, is a congruence on **A**, if and only if it is the kernel, ker(α), of some homomorphism α:**A**$\to$**B**, for some algebra **B**. Also the homomorphic image $\alpha(A)$ is isomorphic to $A/\ker(\alpha)$ via the isomorphism which sends $\alpha(a)$ to $a\cdot\ker(\alpha)$ for every element $a\in A$.

Let q and τ be congruences on A. Define $q\leq\tau$ to mean $q\subseteq\tau$ (as subsets of $A\times A$) or equivalently, aqb implies $a\tau b$ for all elements $a,b\in A$. The partial order $\leq$ on all congruences on **A** induces a complete lattice (see section 2). The meet of congruences is the intersection of congruences (as subsets of $A\times A$) and the join of congruences is the smallest congruence containing their union.

An interesting property, possessed by every unary variety $\mathscr{V}$ is the so called *congruence extension property* [5]. This says that every congruence τ on a subalgebra B of unary algebra $A\in\mathscr{V}$, can be extended to all of **A**. That is, there exists a congruence q on A such that aqb exactly when $a\tau b$, for all elements $a,b\in B$. For example if $i=<\{(a,a)\,|\,a\in A_s\}\,|\,s\in\mathscr{Q}>$, which is trivially a congruence on **A**, then one possibility is $q=\tau\cup i$. An equivalent statement of the congruence extension property is that any subalgebra of any homomorphic image of any algebra $A\in\mathscr{V}$ is a homomorphic image of a subalgebra of **A**. (The converse is true for all varieties.)

Some congruences on unary algebras can be obtained in the following way. Let **A** be a unary algebra and G a group of automorphisms of **A**. Define a relation q on A where for all elements $a,b\in A$ it holds that aqb exactly when $a=\alpha(b)$ for some automorphism $\alpha\in G$. Since G is a group, q is an equivalence. Also q is $\mathscr{Q}$-respecting, since each automorphism $\alpha\in G$ is. For any sort $s\in\mathscr{Q}$ and any elements $a,b\in A$, if aqb then there exists automorphism $\alpha\in G$ such that $a=\alpha(b)$, by definition. Since α is a homomorphism it follows that $f_{\mathbf{A}}(a)=f_{\mathbf{A}}(\alpha(b))=\alpha(f_{\mathbf{A}}(b))$ for every sort $t\in\mathscr{Q}$ and every operator symbol $f\in\mathscr{O}_{s,t}$. Therefore q is a congruence. Not all congruences can be obtained in this way.

We now illustrate some of the above concepts with familiar examples of minor classes. If $g:B\to D$ is a function then there is a (minor class) homomorphism from $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ to $\mathscr{F}^{\mathscr{Q}}(K,C,D)$ which sends each $f:A^{Q}\to B$

(where $A=K\cup C$) to $(g\circ f):A^Q\to D$. (For non–injective $g:B\to D$, the effect of this homomorphism is like looking at the "coloured grid" for f with a partial colour blindness, in the visualisation given in section 4.) The homomorphism is an isomorphism if g is a bijection, and an automorphism if $B=D$ and g is a permutation of B. If $E\subseteq C$, then there is a homomorphism from $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ to $\mathscr{F}^{\mathscr{Q}}(K,E,B)$ which sends each $f:A^Q\to B$ to the restriction of f to $(K\cup E)^Q$. (This amounts to discarding some of the "coloured grid" for f.)

The minor class $\mathscr{W}^{\mathscr{Q}}$ is isomorphic to $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,B)$ where $K=\{$*delete, contract*$\}$ and $|B|=2$, say $B=\{$*in,out*$\}$. For every ground set $Q\in\mathscr{Q}$ and every $P\subseteq Q$ let $\overline{P}:Q\to K$ be the function with $\overline{P}(q)=$*delete* whenever $q\notin P$ and $\overline{P}(q)=$*contract* whenever $q\in P$. This means that K^Q is $\{\overline{P}\,|\,P\subseteq Q\}$. For every $W\subseteq 2^Q$ let $\overline{W}:K^Q\to B$ be the function with $\overline{W}(\overline{P})=$*in* whenever $P\in W$ and $\overline{W}(\overline{P})=$*out* whenever $P\notin W$. This means that $B^{(K^Q)}$ is $\{\overline{W}\,|\,W\subseteq 2^Q\}$. The function from $\mathscr{W}^{\mathscr{Q}}$ to $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,B)$ which sends (W,Q) to $(\overline{W},Q)$ is an isomorphism. So we can visualise (W,Q) as the points of a $2\times 2\times\ldots\times 2$ ($|Q|$ times) grid, or equivalently, the vertices of a $|Q|$–dimensional hypercube, each with one of two possible labels, like in/out, black/white, etcetera. Any automorphism of this particular $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,B)$ corresponds to an automorphism of $\mathscr{W}^{\mathscr{Q}}$. From the previous paragraph it follows that, the identity function $\mathscr{W}^{\mathscr{Q}}$ and the function sending each $(W,Q)\in\mathscr{W}^{\mathscr{Q}}$ to $(2^Q-W,Q)$, are both automorphisms of $\mathscr{W}^{\mathscr{Q}}$. The second of these comes from swapping the two elements of B for each function $f:K^Q\to B$, or for $W\subseteq 2^Q$, swapping membership and non–membership of 2^Q, giving 2^Q-W, the complement of W. These two automorphisms form a group so there is a congruence q on $\mathscr{W}^{\mathscr{Q}}$ where $V\mathsf{q}W$ if and only if $G(V)=G(W)$ (let these equal Q), and either $V=W$ or $V=2^Q-W$. This defines the minor class $\mathscr{W}^{\mathscr{Q}}/\mathsf{q}$, and in particular $\mathscr{W}/\mathsf{q}$ (where the ground sets are the finite sets). For any set Q, a *clutter* on ground set Q is a structure (W,Q), where $W \subseteq 2^Q$ and for any P,R∈W, if P⊆R then P=R. The clutters with finite ground sets form a subset of $\mathscr{W}$ which is actually a sub minor class of $\mathscr{W}$, (since it is closed under deletion, contraction and clutter isomorphism). However, Seymour [14] defines a different minor class whose structures are clutters, and this is not a sub minor class of $\mathscr{W}$. However it is isomorphic to $\mathscr{W}^U$ (defined earlier in this section) via the isomorphism which sends each clutter W, on ground set Q, to the structure $\{P \mid R \subseteq P \subseteq Q$ for some $R \in W\}$, on ground set Q, in $\mathscr{W}^U$. Since $\mathscr{W}^U$ is a sub minor class of $\mathscr{W}$, Seymour's minor class of clutters can be embedded in $\mathscr{W}$.

Direct Products

Let $\Sigma=(\mathscr{Q},\mathscr{O})$ be a unary signature. If I is a set and $(\mathbf{A}^i \mid i \in I)$ is a family of Σ-algebras then the algebra $\mathbf{P}=\prod_{i\in I} \mathbf{A}^i$, defined below, is the *direct product* of $(\mathbf{A}^i \mid i \in I)$. The universe of **P** is $P=\prod_{i\in I} \langle A^i_s \mid s \in \mathscr{Q} \rangle$ (see section 2). That is, for any sort $s \in \mathscr{Q}$, the elements of P_s are of the form $(a^i \mid i \in I)$ where $a^i \in A^i_s$ for all $i \in I$. (The components of $(a^i \mid i \in I)$ all have the <u>same</u> sort.) For all sorts $s,t \in \mathscr{Q}$, every operator symbol $f \in \mathscr{O}_{s,t}$, and every element $(a^i \mid i \in I) \in P_s$ it holds that $f_{\mathbf{P}}((a^i \mid i \in I))=(f_{\mathbf{A}^i}(a^i) \mid i \in I)$. In particular, if $I=\emptyset$ then $|P_s|=1$ for every sort $s \in \mathscr{Q}$ and $f_{\mathbf{P}}:P_s \to P_t$ is defined in the only possible way for every $f \in \mathscr{O}_{s,t}$ and all $s,t \in \mathscr{Q}$.

For example, let I be a set and $(\mathscr{S}^i \mid i \in I)$ a family of $(\mathscr{Q},K)$ minor classes (for some set of ground sets $\mathscr{Q}$ and set of manners K). Then $\prod_{i\in I} \mathscr{S}^i$ has structures of the form $((S^i \mid i \in I),Q)$ where $Q \in \mathscr{Q}$ and $S^i \in \mathscr{S}^i_Q$ for every $i \in I$. That is, the structures in $(S^i \mid i \in I)$ all have the same ground set, Q. For every appropriate prescription $\mathfrak{K}$ and bijection ω it holds that $(S^i \mid i \in I)[\mathfrak{K}]$ is $(S^i[\mathfrak{K}] \mid i \in I)$ and $\omega((S^i \mid i \in I))$ is $(\omega(S^i) \mid i \in I)$.

Suppose I is a set, $(B^i \mid i \in I)$ is a family of sets, and let $B = \prod_{i \in I} B^i$. Then $\prod_{i \in I} \mathscr{F}^{\mathscr{Q}}(K,C,B^i)$ is isomorphic to $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ via the isomorphism which sends each $(f^i \mid i \in I)$ on ground set Q, (so that if $f^i : A^Q \to B^i$ for each $i \in I$) to the function $f : A^Q \to B$, on ground set Q in $\mathscr{F}^{\mathscr{Q}}(K,C,B)$, where $f(x) = (f^i(x) \mid i \in I)$ for every $x \in A^Q$. If $|I| = 2$ and we use our coloured grid visualisation, then $\prod_{i \in I} \mathscr{F}^{\mathscr{Q}}(K,C,B^i)$ consists of pairs of coloured grids (with the same ground sets), while the elements of $\mathscr{F}^{\mathscr{Q}}(K,C,B)$ are single grids where each gridpoint is labelled with an ordered pair of colours, as though the two grids were superimposed. If, for example, one grid was coloured black and white, and the other, red and blue, then in the resulting single grid we could "mix" the two colours at each gridpoint, to colour it dark red, light red, dark blue or light blue.

SECTION 6: CONSTRUCTIONS WHICH MODIFY THE SIGNATURE

Usually in universal algebra the signature Σ is kept fixed throughout some discussion. However there exist in the theory, constructions to obtain a Σ'–algebra from a Σ–algebra, where Σ' may be different from Σ but is related to it in some way. Three such constructions are worth mentioning for minor classes.

For a minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K,\mathcal{P})$ (abbreviated $\mathcal{S}$) and hereditary $\mathcal{Q}'\subseteq\mathcal{Q}$, the minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K,\mathcal{P})|\mathcal{Q}'$ (abbreviated $\mathcal{S}|\mathcal{Q}'$) is called $\mathcal{S}$ *restricted to* $\mathcal{Q}'$. It is the minor class $(\mathcal{S}|\mathcal{Q}',\mathcal{Q}',\mathcal{I}|\mathcal{Q}',K,\mathcal{P}|\mathcal{Q}')$ where $\mathcal{S}|\mathcal{Q}'=\{(S,Q)|(S,Q)\in\mathcal{S},Q\in\mathcal{Q}'\}$ contains only those structures in $\mathcal{S}$ with ground sets in $\mathcal{Q}'$, and $\mathcal{I}|\mathcal{Q}'$ and $\mathcal{P}|\mathcal{Q}'$ contain only those functions in $\mathcal{I}$ and $\mathcal{P}$ (respectively) which involve only structures in $\mathcal{S}|\mathcal{Q}'$. For example $\mathcal{F}^{\mathcal{Q}}(K,C,B)|\mathcal{Q}'$ is $\mathcal{F}^{\mathcal{Q}'}(K,C,B)$. The corresponding signature $\Sigma=(\mathcal{Q},\mathcal{O})$ is modified by removing all $Q\in\mathcal{Q}-\mathcal{Q}'$ from $\mathcal{Q}$ and removing all operator symbols involving these sorts from $\mathcal{O}$.

The minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K,\mathcal{P})|K'$ (abbreviated $\mathcal{S}|K'$ and called $\mathcal{S}$ *confined to* K') where $K'\subseteq K$ is the minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K',\mathcal{P}|K')$ where $\mathcal{P}|K'$ contains only those point removals involving only manners of point removal in K'. For example if $\mathcal{M}$ is the minor class of matroids with deletion and contraction, then $\mathcal{M}|\{\text{delete}\}$ is the minor class of matroids with deletion only. The signature $\Sigma=(\mathcal{Q},\mathcal{O})$ is modified by removing from $\mathcal{Q}$, all point removals involving manners of point removal in $K-K'$.

For a minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K,\mathcal{P})$ and a bijection $\kappa:K\to K'$ the minor class $\kappa(\mathcal{S},\mathcal{Q},\mathcal{I},K,\mathcal{P})$ (abbreviated $\kappa(\mathcal{S})$) is the minor class $(\mathcal{S},\mathcal{Q},\mathcal{I},K',\kappa(\mathcal{P}))$ where $\kappa(\mathcal{P})$ is obtained from $\mathcal{P}$ by renaming "point removal in manner ℓ" as "point removal in manner $\kappa(\ell)$" for every $\ell\in K$. Similarly, the signature $\Sigma=(\mathcal{Q},\mathcal{O})$ is modified by changing each mention of ℓ to $\kappa(\ell)$ for every $\ell\in K$.

For example if κ swaps deletion and contraction then $\kappa(\mathscr{M})$ is the minor class of matroids with the usual meaning of deletion and contraction swapped.

A *mixed homomorphism* from $(\mathscr{S},\mathscr{Q},\mathscr{T},\mathrm{K},\mathscr{P})$ to $(\mathscr{S}',\mathscr{Q},\mathscr{T}',\mathrm{K}',\mathscr{P}')$ (notice $\mathscr{Q}$ is the same in both minor classes) consists of a bijection $\kappa{:}\mathrm{K}\to\mathrm{K}'$ and a (conventional) homomorphism from $(\mathscr{S},\mathscr{Q},\mathscr{T},\mathrm{K}',\kappa(\mathscr{P}))$ to $(\mathscr{S}',\mathscr{Q},\mathscr{T}'\mathrm{K}',\mathscr{P}')$. Mixed isomorphism and mixed automorphism are defined analogously. For example, while the minor class of matroids $\mathscr{M}$ has only one automorphism (the identity), which can also be treated as a mixed automorphism, it has one other mixed automorphism, namely the one which swaps deletion and contraction and which sends each matroid to its dual.

SECTION 7: EQUIVALENT CHARACTERISATIONS OF VARIETIES

Let $\Sigma=(\mathcal{S},\mathcal{O})$ be a unary signature and let $Y=\langle Y_s \mid s\in\mathcal{S}\rangle$ be a set of variables. Recall the definition of $T^{\Sigma}(Y)=\langle T^{\Sigma}_s(Y) \mid s\in\mathcal{S}\rangle$ where $T^{\Sigma}_s(Y)$ is the set of Σ-terms of sort s in variables Y. (See section 5.) The Σ-algebra $\mathbf{T}^{\Sigma}(Y)$, with universe $T^{\Sigma}(Y)$ is defined in a natural way [1,6]. That is, for all sorts $s,t\in\mathcal{S}$, and every operator symbol $f\in\mathcal{O}_{s,t}$, the function $f_{\mathbf{T}^{\Sigma}(Y)}:T^{\Sigma}_s(Y)\to T^{\Sigma}_t(Y)$ sends each term of sort s, $\ell\in T^{\Sigma}_s(Y)$, to the term of sort t, $f\ell\in T^{\Sigma}_t(Y)$. (The term $f\ell$ is a string of symbols obtained by simply appending the symbol f to the left hand end of ℓ.)

For any Σ-algebra **A** and any assignment $\alpha:Y\to A$, its extension $\overline{\alpha}:\mathbf{T}^{\Sigma}(Y)\to\mathbf{A}$ is actually the unique homomorphism such that $\overline{\alpha}(y)=\alpha(y)$ for every $y\in Y$. (Note that Y generates $\mathbf{T}^{\Sigma}(Y)$.) Also, if $\mathscr{C}$ is a class of Σ-algebras, then the $\mathcal{S}$-respecting equivalence relation $eq^{Y}(\mathscr{C})$ is in fact a congruence of $\mathbf{T}^{\Sigma}(Y)$. Moreover, $eq^{Y}(\mathscr{C})$ is the meet of congruences $\ker(\overline{\alpha})$ for all assignments $\alpha:Y\to A$ and all algebras $\mathbf{A}\in\mathscr{C}$. The significance of the Σ-algebra $\mathbf{T}^{\Sigma}(Y)/eq^{Y}(\mathscr{C})$ becomes apparent below.

Let **F** be a Σ-algebra and let $\alpha:Y\to F$ be an assignment whose extension $\overline{\alpha}:\mathbf{T}^{\Sigma}(Y)\to\mathbf{F}$ is a surjective homomorphism. The pair $(\mathbf{F},\alpha)$ is *universal* for a class of Σ-algebras $\mathscr{C}$, if for every algebra $\mathbf{A}\in\mathscr{C}$ and every assignment $\gamma:Y\to A$ (with unique extension $\overline{\gamma}:\mathbf{T}^{\Sigma}(Y)\to\mathbf{A}$) there exists (uniquely, it turns out) a homomorphism $\beta:\mathbf{F}\to\mathbf{A}$ such that $\overline{\gamma}$ is $\beta\circ\overline{\alpha}$. (For example, $(\mathbf{T}^{\Sigma}(Y),1)$, where $1:Y\to Y$ is the identity function, is universal for $\mathscr{C}$.) In general, such a homomorphism β exists exactly when $\ker(\overline{\alpha})\leq\ker(\overline{\gamma})$ (in the congruence ordering). Now $\ker(\overline{\alpha})$ is less than all possible $\ker(\overline{\gamma})$ exactly when it is less than their meet, which is $eq^{Y}(\mathscr{C})$. Therefore, $(\mathbf{F},\alpha)$ is universal for $\mathscr{C}$, if and only if $\overline{\alpha}:\mathbf{T}^{\Sigma}(Y)\to\mathbf{F}$ is surjective (so that $\mathbf{F}\cong\mathbf{T}^{\Sigma}(Y)/\ker(\overline{\alpha})$) and $\ker(\overline{\alpha})\leq eq^{Y}(\mathscr{C})$. In particular, $\ker(\overline{\alpha})$ can be chosen to be maximal, that is, $\ker(\overline{\alpha})=eq^{Y}(\mathscr{C})$, as follows. Let $F=T^{\Sigma}(Y)/eq^{Y}(\mathscr{C})$ and define $\overline{\alpha}:T^{\Sigma}(Y)\to F$ in the natural way, that is $\alpha(\ell)=\ell\cdot eq^{Y}(\mathscr{C})$ for every term $\ell\in T^{\Sigma}(Y)$.

For convenience, $T^{\Sigma}(Y)/eq^{Y}(\mathscr{C})$ is abbreviated to $F^{\mathscr{C}}(Y)$, or $F(Y)$ when $\mathscr{C}$ is known from the context.

For a class $\mathscr{C}$ of Σ-algebras let $S(\mathscr{C})$ (respectively $H(\mathscr{C}),P(\mathscr{C})$) be the class of subalgebras (respectively homomorphic images, direct products) of algebras in $\mathscr{C}$. If $\mathscr{C}=S(\mathscr{C})$ (respectively $\mathscr{C}=H(\mathscr{C})$, $\mathscr{C}=P(\mathscr{C})$) then $\mathscr{C}$ is closed under subalgebras (respectively homomorphic images, direct products). Let $V(\mathscr{C})$ be the smallest variety containing $\mathscr{C}$, namely $alg(eq(\mathscr{C}))$.

A fundamental result in universal algebra [1,6], says that the following are equivalent.

(1) $\mathscr{C}$ is a variety.

(2) $\mathscr{C}$ is closed under subalgebras, homomorphic images and direct products.

(3) $\mathscr{C}$ is closed under homomorphic images and contains $F^{\mathscr{C}}(Y)$ for every set of variables Y.

In particular, a variety of minor classes is closed under sub minor classes, homomorphic images and direct products.

Define the *support* of a Σ-algebra **A**, denoted supp(**A**), to be the set of sorts $s\in\mathscr{S}$ such that A_s is non-empty; that is, $supp(A)=\{s\,|\,s\in\mathscr{S}$ and $A_s\neq\emptyset\}$. If $supp(A)=\mathscr{S}$, then **A** *has full support*. It is routine to show that the following are equivalent.

(1′) $\mathscr{C}$ is a proper variety.

(2′) $\mathscr{C}$ is closed under subalgebras, homomorphic images and direct products, and every algebra in $\mathscr{C}$, has an extension in $\mathscr{C}$, which has full support. (That is, for all $A\in\mathscr{C}$, there exists $B\in\mathscr{C}$ such that $A\leq B$ and $supp(B)=\mathscr{S}$.)

(3′) $\mathscr{V}$ is closed under homomorphic images, and contains $\mathbf{F}^{\mathscr{V}}(Y)=\mathbf{T}^{\Sigma}(Y)/eq^{Y}(\mathscr{V})$ for every set of variables Y, and for $Z\subseteq Y$, $eq^{Z}(\mathscr{V})$ contains precisely those pairs (ℓ,r) such that $(\ell,r)\in eq^{Y}(\mathscr{V})$ and $var(\ell,r)\subseteq Z$,

Note that, in an improper variety $\mathscr{V}$, $eq^{Z}(\mathscr{V})$ need not contain all these pairs.

The above results hold for all algebras. We now confine our attention to unary algebras. Let I be a set, let **A** be a Σ-algebra, and let $(\mathbf{A}^{i}|i\in I)$ be a family of subalgebras of **A**. If $A=\bigcup_{i\in I} A^{i}$, then **A** is *the union of* $(\mathbf{A}^{i}|i\in I)$. If, in addition, the members of $(\mathbf{A}^{i}|i\in I)$ are pairwise disjoint, then **A** is *the disjoint union of* $(\mathbf{A}^{i}|i\in I)$. More generally, if $(\mathbf{B}^{i}|i\in I)$ is a family of Σ-algebras with $\mathbf{B}^{i}\cong\mathbf{A}^{i}$, for every $i\in I$, then **A** is *a disjoint union of* $(\mathbf{B}^{i}|i\in I)$, and again, if the members of $(\mathbf{A}^{i}|i\in I)$ are pairwise disjoint, then **A** is *a disjoint union of* $(\mathbf{B}^{i}|i\in I)$. For any family $(\mathbf{B}^{i}|i\in I)$ of Σ-algebras, there exists a Σ-algebra **A** which is a disjoint union of $(\mathbf{B}^{i}|i\in I)$, since we can take isomorphic copies $\mathbf{A}^{i}\cong\mathbf{B}^{i}$ with the members of $(\mathbf{A}^{i}|i\in I)$ pairwise disjoint, put $A=\bigcup_{i\in I} A^{i}$, and let **A** be the unique Σ-algebra, with universe A, such that $\mathbf{A}^{i}=\mathbf{A}incA^{i}$ for every $i\in I$. Since a disjoint union of $(\mathbf{B}^{i}|i\in I)$ is unique up to isomorphism, it is called <u>the</u> disjoint union of $(\mathbf{B}^{i}|i\in I)$. Observe that any union of $(\mathbf{B}^{i}|i\in I)$ is a homomorphic image of the disjoint union of $(\mathbf{B}^{i}|i\in I)$.

For a class $\mathscr{V}$ of Σ-algebras, let $D(\mathscr{V})$ be the class of disjoint unions of algebras in $\mathscr{V}$. If $\mathscr{V}=D(\mathscr{V})$, then $\mathscr{V}$ is closed under disjoint unions. A unary variety $\mathscr{V}$ is closed under disjoint unions, if and only if it has the so called *amalgamation property* [5]. It is routine to show that the following are equivalent.

(1″) $\mathscr{V}$ is a regular (unary) variety.

(2″) $\mathscr{V}$ is closed under subalgebras, homomorphic images, direct products and disjoint unions.

(3″) $\mathscr{C}$ is closed under homomorphic images, and contains $\mathbf{F}^{\mathscr{C}}(\mathrm{Y})$ for every set of variables Y, and also $\mathbf{F}^{\mathscr{C}}(\mathrm{Y})$ is the disjoint union of $(\mathbf{F}^{\mathscr{C}}(\{y\}) \mid y \in \mathrm{Y})$.

In particular, if $\mathscr{C}$ is the class of $(\mathscr{Q},\mathrm{K})$ minor classes (for some set of ground sets $\mathscr{Q}$ and set of manners K), then $\mathscr{C}$ is a regular unary variety, and all of the nine statements (1) to (3″) hold.

Let $\mathscr{C}$ be a regular unary variety of Σ–algebras, let algebra $\mathbf{F}\in\mathscr{C}$, and let $\mathrm{G}\subseteq\mathrm{F}$. Let $1{:}\mathrm{G}\to\mathrm{F}$ be the identity assignment and let $\bar{1}{:}\mathrm{T}^{\Sigma}(\mathrm{G})\to\mathrm{F}$ be its extension (which is surjective, if and only if **F** is generated by G). If $\bar{1}$ is surjective and $\ker(\bar{1})=\mathrm{eq}^{\mathrm{Y}}(\mathscr{C})$, then **F** is *a free algebra of* $\mathscr{C}$, *freely generated by* G. In particular, $\mathbf{F}^{\mathscr{C}}(\mathrm{Y})$ is a free algebra of $\mathscr{C}$, freely generated by Y (or strictly speaking by $\mathrm{Y}/\mathrm{eq}^{\mathrm{Y}}(\mathscr{C})$ which can be harmlessly identified with Y). Every free algebra of $\mathscr{C}$, freely generated by Y, is isomorphic to $\mathbf{F}^{\mathscr{C}}(\mathrm{Y})$. By (3″), discussion of free algebras can usually be restricted to free algebras freely generated by a single element.

A useful result is that for any class $\mathscr{C}$ of Σ–algebras, $\mathrm{V}(\mathscr{C})=\mathrm{H}(\mathrm{S}(\mathrm{P}(\mathscr{C})))$. This is proved for the 1–sorted case ($|\mathscr{Q}|=1$) in [5], but this generalises easily to the many–sorted case. If Σ is a unary signature then, as noted earlier, any subalgebra of a homomorphic image of a Σ–algebra **A**, is a homomorphic image of a subalgebra of **A**, and visa–versa. So in the unary case, $\mathrm{V}(\mathscr{C})$ also equals $\mathrm{S}(\mathrm{H}(\mathrm{P}(\mathscr{C})))$, a result used later.

SECTION 8: TWO EMBEDDING THEOREMS

Often in algebra the question is asked, "When can every algebra in a variety be embedded in an algebra of a certain type ?" An example is Cayley's theorem which states that any group can be embedded in the group of permutations of some set. Although far from unique for any given variety, such theorems tell us something about what the algebras in the variety "look like". The first embedding theorem (8.1) states that any minor class can be embedded in a minor class of the form $\mathcal{F}^{\mathcal{Q}}(K,C,B)$, so the structures of any minor class can be visualised as "coloured grids" (see section 4). The second embedding theorem (8.4) states that any minor class can be embedded in a minor class of the form $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)/\mathfrak{q}$, for some congruence $\mathfrak{q}$.

<u>Theorem 8.1</u>: If $(\mathcal{S},\mathcal{Q},\mathcal{T},K,\mathcal{P})$ is a minor class, then there exist sets C and B such that $(\mathcal{S},\mathcal{Q},\mathcal{T},K,\mathcal{P})$ can be embedded in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$.

Note that in the above theorem, $\mathcal{S},\mathcal{Q},\mathcal{T}$,K and $\mathcal{P}$ are sets (as stated in section 5). The theorem would still hold if they were classes, provided C and B were allowed to be classes. Two lemmas are given before the proof of theorem 8.1.

<u>Lemma 8.2</u>: Let $(\mathcal{S},\mathcal{Q},\mathcal{T})$ and $(\mathcal{T},\mathcal{Q},\mathcal{J})$ be isomorphism classes and let the ($\mathcal{Q}$-partitioned) set of structures $\mathcal{S}^*$ be such that $\mathcal{S}^*\subseteq\mathcal{S}$ and every structure in $\mathcal{S}$ is isomorphic to exactly one structure in $\mathcal{S}^*$. That is, if $\mathcal{S}$ is partitioned into equivalence classes, where the equivalence is isomorphism, then $\mathcal{S}^*$ contains exactly one representative from each equivalence class. (Note that $\mathcal{S}^*$ is almost never a subuniverse of $\mathcal{S}$.) Let $\alpha:\mathcal{S}^*\to\mathcal{T}$ be a $\mathcal{Q}$-respecting function. Then the following are equivalent.

(i) There exists an embedding $\overline{\alpha}:\mathscr{S}\to\mathscr{T}$ such that $\overline{\alpha}(S)=\alpha(S)$ for every structure $S\in\mathscr{S}^*$. (Note that if $\overline{\alpha}$ exists, then it exists uniquely, since $\mathscr{S}^*$ generates $\mathscr{S}$. See section 5.)

(ii) If structure $S\in\mathscr{S}^*$ then $\mathrm{Aut}(S)=\mathrm{Aut}(\alpha(S))$, and if structures $S,T\in\mathscr{S}^*$ and $S\neq T$ then $\alpha(S)\ncong\alpha(T)$.

(Note that $S\in\mathscr{S}^*$ implies that $S\in\mathscr{S}$ so that Aut(S) is well defined.)

Proof: Suppose that $\overline{\alpha}$ exists and is an embedding. Let structure $S\in\mathscr{S}^*$ and let $\omega:G(S)\to G(S)$ be a bijection. Then $S=\omega(S)$ exactly when $\overline{\alpha}(S)=\overline{\alpha}(\omega(S))$, since $\overline{\alpha}$ is injective, and $\overline{\alpha}(\omega(S))=\omega(\overline{\alpha}(S))$, since $\overline{\alpha}$ is a homomorphism. It follows that $\mathrm{Aut}(S)=\mathrm{Aut}(\overline{\alpha}(S))=\mathrm{Aut}(\alpha(S))$ for every structure $S\in\mathscr{S}^*$.

Let S and T be structures in $\mathscr{S}^*$ and suppose that $\alpha(S)\cong\alpha(T)$. Then $\overline{\alpha}(S)\cong\overline{\alpha}(T)$, so that $\overline{\alpha}(S)=\omega(\overline{\alpha}(T))$ for some bijection $\omega:G(\overline{\alpha}(T))\to G(\overline{\alpha}(S))$. Since $\overline{\alpha}$ is a homomorphism, $\overline{\alpha}(S)=\omega(\overline{\alpha}(T))=\overline{\alpha}(\omega(T))$, and since $\overline{\alpha}$ is injective, $S=\omega(T)$. Thus $S\cong T$, which implies $S=T$ by the definition of $\mathscr{S}^*$. Equivalently, if $S\neq T$, then $\alpha(S)\ncong\alpha(T)$. Therefore (i) implies (ii).

Conversely, suppose the conditions of (ii) hold. Every structure $T\in\mathscr{S}$ is isomorphic to a structure $S\in\mathscr{S}^*$, and is therefore of the form $\omega(S)$ for some bijection $\omega:G(S)\to G(T)$. In this expression, S is uniquely determined, as the definition of $\mathscr{S}^*$ guarantees, but ω need not be. (For example, if T=S, then ω could be any automorphism of S.) Define $\overline{\alpha}(\omega(S))$ to be $\omega(\alpha(S))$ for all $\omega(S)\in\mathscr{S}$ (where $S\in\mathscr{S}^*$). Since ω may not be uniquely determined, this might assign more than one value to $\overline{\alpha}(\omega(S))$, for some $\omega(S)\in\mathscr{S}$, and this is the only way $\overline{\alpha}$ could fail to be well defined. Suppose that the bijection $\tau:G(S)\to G(\omega(S))$ is such that $\omega(S)=\tau(S)$. Then $(\tau^{-1}\circ\omega)(S)=(S)$, so that $(\tau^{-1}\circ\omega)\in\mathrm{Aut}(S)$. But then $(\tau^{-1}\circ\omega)\in\mathrm{Aut}(\alpha(S))$, since $\mathrm{Aut}(S)\subseteq\mathrm{Aut}(\alpha(S))$ by (ii). Thus $(\tau^{-1}\circ\omega)(\alpha(S))=\alpha(S)$ so that $\omega(\alpha(S))=\tau(\alpha(S))$. Therefore $\overline{\alpha}(\omega(S))$ is well defined and $\overline{\alpha}$ exists. In particular, if $S\in\mathscr{S}^*$ and $1:G(S)\to G(S)$ is the identity function, then $\overline{\alpha}(S)=\overline{\alpha}(1(S))=1(\alpha(S))=\alpha(S)$, as required.

Also $\overline{\alpha}$ is a homomorphism as the following argument shows. Let $\omega(S)\in\mathscr{S}$ (where $S\in\mathscr{S}^*$) and let τ be a (structure) isomorphism acting on $\omega(S)$. Then using the definition of $\overline{\alpha}$ (twice) it follows that $\overline{\alpha}(\tau(\omega(S)))=\overline{\alpha}((\tau\circ\omega)(S))=(\tau\circ\omega)(\alpha(S))=\tau(\omega(\alpha(S)))=\tau(\overline{\alpha}(\omega(S)))$, as required. In fact $\overline{\alpha}$ is the unique homomorphism agreeing with α on $\mathscr{S}^*$ (since $\mathscr{S}^*$ generates $\mathscr{S}$).

We now show that $\overline{\alpha}$ is injective. Let the structures $\omega(S),\tau(T)\in\mathscr{S}$ (where $S,T\in\mathscr{S}^*$) be such that $\overline{\alpha}(\omega(S))=\overline{\alpha}(\tau(T))$. By the definition of $\overline{\alpha}$, $\omega(\alpha(S))=\tau(\alpha(\tau))$ and hence $(\tau^{-1}\circ\omega)(\alpha(S))=\alpha(T)$. Thus $\alpha(S)\cong\alpha(T)$ and by (ii), $S=T$. Therefore $(\tau^{-1}\circ\omega)(\alpha(S))=\alpha(S)$ which implies $(\tau^{-1}\circ\omega)\in\mathrm{Aut}(\alpha(S))$, and also $(\tau^{-1}\circ\omega)\in\mathrm{Aut}(S)$, since $\mathrm{Aut}(\alpha(S))\subseteq\mathrm{Aut}(S)$, by (ii). Hence $(\tau^{-1}\circ\omega)(S)=(S)$, so that $\omega(S)=\tau(S)=\tau(T)$. It follows that the homomorphism $\overline{\alpha}$ is injective, and hence an embedding, as required. ☺

Observe that a minor class $(\mathscr{S},\mathscr{Q},\mathscr{T}\ \emptyset,\emptyset)$ with no manners of point removal is, for all intents and purposes, the isomorphism class $(\mathscr{S},\mathscr{Q},\mathscr{T})$. In particular, $\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ is an isomorphism class. Then next lemma is a special case of theorem 8.1 where $K=\emptyset$.

<u>Lemma 8.3</u>: Let $(\mathscr{S},\mathscr{Q},\mathscr{T})$ be an isomorphism class. Then there exist sets C and B such that $(\mathscr{S},\mathscr{Q},\mathscr{T})$ can be embedded in $\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$.

<u>Proof</u>: Let $\mathscr{S}^*$ be defined as in lemma 8.2. To prove lemma 8.3, it is sufficient to construct a $\mathscr{Q}$-respecting function $\alpha:\mathscr{S}^*\to\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ satisfying condition (ii) of lemma 8.2. (For each structure $S\in\mathscr{S}^*$, we construct a "pattern" on the "coloured grid" $\alpha(S)\in\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$, with the appropriate automorphisms, and with α sending distinct elements of $\mathscr{S}^*$, to non isomorphic "coloured grids".)

Let B_1 and B_2 be disjoint sets with the same cardinality as $\mathscr{S}^*$, let $B=B_1\cup B_2$, and let $g_1:\mathscr{S}^*\to B_1$ and $g_2:\mathscr{S}^*\to B_2$ be bijections. Let C be a set with larger cardinality than that of any ground set in $\mathscr{Q}$, and for each structure $S\in\mathscr{S}^*$, let $h_S:G(S)\to C$ be an injection. If $G(S)=Q$, define $Sym(S)\subseteq C^Q$ to be $\{h_S\circ\omega\,|\,\omega\in Aut(S)\}$.

Let $\alpha:\mathscr{S}^*\to\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ be defined as follows. For any structure $S\in\mathscr{S}^*$, with ground set $G(S)=Q$, the function $\alpha(S):C^Q\to B$ is given by $(\alpha(S))(x)=g_1(S)$ whenever $x\in Sym(S)$, and $(\alpha(S))(x)=g_2(S)$ whenever $x\in C^Q-Sym(S)$. Recall that, by definition, $(\tau(\alpha(S)))(x)=(\alpha(S))(\tau^{-1}(x))$ and $\tau^{-1}(x)=x\circ\tau$ for every $x\in C^Q$. (See section 3.) It follows that

$Aut(\alpha(S))$

$=\{\tau\,|\,(\tau:Q\to Q \text{ is a bijection}) \text{ and } \tau(\alpha(S))=\alpha(S)\}$

$=\{\tau\,|\,(\alpha(S))(x\circ\tau)=\alpha(S)(x) \text{ for every } x\in C^Q\}$

$=\{\tau\,|\,(x\circ\tau)\in Sym(S) \text{ exactly when } x\in Sym(S) \text{ for every } x\in C^Q\}$

$=(\tau\,|\,\{h_S\circ\omega\circ\tau\,|\,\omega\in Aut(S)\}=Sym(S)\}$

If $\tau\in Aut(S)$, then $\omega\circ\tau\in Aut(S)$ exactly when $\omega\in Aut(S)$, so that $\{h_S\circ\omega\circ\tau\,|\,\omega\in Aut(S)\}=\{h_S\circ\omega\circ\tau\,|\,\omega\circ\tau\in Aut(S)\}=Sym(S)$ and hence $\tau\in Aut(\alpha(S))$. Conversely, if $\tau\notin Aut(S)$, then $\omega\circ\tau$ is not the identity function for any $\omega\in Aut(S)$, and since h_S is injective, $h_S\notin\{h_S\circ\omega\circ\tau\,|\,\omega\in Aut(S)\}$, whereas $h_S\in Sym(S)$, so that $\tau\notin Aut(\alpha(S))$. Therefore $Aut(\alpha(S))=Aut(S)$ for all $S\in\mathscr{S}^*$.

Finally, suppose the structures $S,T\in\mathscr{S}^*$ are such that $\alpha(S)\cong\alpha(T)$. The function $\alpha(S):C^{G(S)}\to B$ only takes the values $g_1(S)\in B_1$ and $g_2(S)\in B_2$ while the function $\alpha(T):C^{G(T)}\to B$ only takes the values $g_1(T)\in B_1$ and $g_2(T)\in B_2$. If $\mathscr{S}$ is empty, then the lemma holds trivially, and otherwise C is clearly non-empty. Since $\alpha(S)$ and $\alpha(T)$ are isomorphic, and $B_1\cap B_2=\emptyset$, it must be that $g_1(S)=g_1(T)$ or $g_2(S)=g_2(T)$. Since g_1 and g_2 are injective it follows that $S=T$, as required.

Therefore $\alpha:\mathscr{S}^*\to\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ satisfies (ii) and the embedding $\overline{\alpha}:\mathscr{S}\to\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ exists, as required. ☺

Proof of theorem 8.1: Let $\overline{\alpha}:\mathscr{S}\to\mathscr{F}^{\mathscr{Q}}(\emptyset,C,B)$ be the (isomorphism class) embedding defined in lemma 8.3. Assume, without loss of generality, that $K\cap C=\emptyset$ and let $A=K\cup C$ and $\overline{K}=K\cup\{\odot\}$ for some $\odot\notin K$, as usual.

For any ground set $Q\in\mathscr{Q}$, the "grid point" $y\in A^Q$ can be expressed uniquely as $\mathfrak{K}\Delta x$ (see section 3) where the prescription $\mathfrak{K}:Q\to\overline{K}$ and "subgrid point" $x:G(\mathfrak{K})\to C$ are defined as follows. For each $q\in Q$, $\mathfrak{K}(q)=y(q)$ whenever $y(q)\in K$, and $\mathfrak{K}(q)=\odot$ whenever $y(q)\in C$. This means that $G(\mathfrak{K})=\{q\,|\,q\in Q \text{ and } \mathfrak{K}(q)=\odot\}=\{q\,|\,q\in Q \text{ and } y(q)\in C\}$. For each $q\in G(\mathfrak{K}),x(q)=y(q)$. The statement $(\mathfrak{K}\Delta x)\in A^Q$ shall mean $y\in A^Q$ and $y=\mathfrak{K}\Delta x$, where y uniquely determines $\mathfrak{K}$ and x (and visa versa, of course).

Define a $\mathscr{Q}$-respecting function $\beta:\mathscr{S}\to\mathscr{F}^{\mathscr{Q}}(K,C,B)$ as follows. For any ground set $Q\in\mathscr{Q}$ and any structure $S\in\mathscr{S}_Q$ let $\beta(S):A^Q\to B$ be the function defined by $(\beta(S))(\mathfrak{K}\Delta x)=(\overline{\alpha}(S[\mathfrak{K}]))(x)$ for every $(\mathfrak{K}\Delta x)\in A^Q$. (Observe that β is indeed $\mathscr{Q}$-respecting.)

Recall again the "coloured grid" visualisation (section 4). The "grid points" in A^Q are partitioned into disjoint "subgrids" of various dimensions. (These subgrids are not the same as those in section 4.) There is a separate subgrid for each prescription $\mathfrak{K}\in\overline{K}^Q$ and the subgrid associated with a particular $\mathfrak{K}$ consists of the gridpoints $\mathfrak{K}\Delta x$ for every $x\in C^{G(\mathfrak{K})}$. This subgrid is "coloured" identically to the "coloured subgrid" $\overline{\alpha}(S[\mathfrak{K}]):C^{G(\mathfrak{K})}\to B$. So the coloured grid $\beta(S):A^Q\to B$ is obtained by "piecing together" in an orderly way, the coloured subgrids $\overline{\alpha}(S[\mathfrak{K}]):C^{G(\mathfrak{K})}\to B$, associated by $\overline{\alpha}$ with each minor $S[\mathfrak{K}]$ of S.

To show that β is an embedding, it is necessary to show that β respects point removals and (structure) isomorphisms (making β a homomorphism) and that it is injective. Respecting point removals follows only from the way the "coloured subgrids" are pieced together. Injectivity follows from the injectivity of (the isomorphism class homomorphism) $\overline{\alpha}$, together with the above "piecing together", and similarly for respecting of (structure) isomorphisms. These facts are proved formally below.

For any ground set $Q \in \mathcal{Q}$, any structure $S \in \mathcal{S}_Q$, any prescription $\mathfrak{J} \in \overline{K}^Q$ and any "gridpoint" $(\mathfrak{K}\Delta x) \in A^{G(\mathfrak{J})}$ it holds that

$(\beta(S[\mathfrak{J}]))(\mathfrak{K}\Delta x)$

$=(\overline{\alpha}((S[\mathfrak{J}])[\mathfrak{K}]))(x)$	(by definition of β)
$=(\overline{\alpha}(S[\mathfrak{J}\Delta\mathfrak{K}]))(x)$	(by condition (M5))
$=(\beta(S))((\mathfrak{J}\Delta\mathfrak{K})\Delta x)$	(by definition of β)
$=(\beta(S))(\mathfrak{J}\Delta(\mathfrak{K}\Delta x))$	(since Δ is associative)
$=((\beta(S))[\mathfrak{J}])(\mathfrak{K}\Delta x)$	(by definition of point removal in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$).

Therefore $\beta(S[\mathfrak{J}])=(\beta(S))[\mathfrak{J}]$, and β respects point removals.

For all ground sets $Q,P \in \mathcal{Q}$, every structure $S \in \mathcal{S}_Q$, every bijection $\omega{:}Q \to P$ and every "gridpoint" $(\mathfrak{K}\Delta x) \in A^Q$, the following holds. First note that every element of A^P is of the form $\omega(\mathfrak{K}\Delta x)$ and, as it is routine to show, $(\omega(\mathfrak{K}\Delta x))=(\omega(\mathfrak{K}))\Delta(\omega|_{G(\mathfrak{K})}(x))$, where $\omega|_{G(\mathfrak{K})}$ is the bijection ω restricted to $G(\mathfrak{K})$. The following chain of equalities hold for the reasons given.

$(\beta(\omega(S)))(\omega(\mathfrak{K}\Delta x))$

$=(\beta(\omega(S)))((\omega(\mathfrak{K}))\Delta(\omega|_{G(\mathfrak{K})}(x)))$	(as stated above)
$=(\overline{\alpha}((\omega(S))[\omega(\mathfrak{K})]))(\omega|_{G(\mathfrak{K})}(x))$	(by definition of β)
$=(\overline{\alpha}(\omega|_{G(\mathfrak{K})}(S[\mathfrak{K}])))(\omega|_{G(\mathfrak{K})}(x))$	(by condition (M3))
$=(\omega|_{G(\mathfrak{K})}(\overline{\alpha}(S[\mathfrak{K}])))(\omega|_{G(\mathfrak{K})}(x))$	(since $\overline{\alpha}$ respects structure isomorphisms)
$=(\overline{\alpha}(S[\mathfrak{K}]))(x)$	(by definition of structure isomorphism in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$)
$=(\beta(S))(\mathfrak{K}\Delta x)$	(by definition of β)
$=(\omega(\beta(S)))(\omega(\mathfrak{K}\Delta x))$	By definition of structure isomorphism in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$).

Therefore $\beta(\omega(S))=\omega(\beta(S))$, and β respects structure isomorphisms. This proves that β is a minor class homomorphism.

Finally, if ground set $Q\in\mathcal{Q}$ and distinct structures $S,T\in\mathcal{S}_Q$, then, since $\overline{\alpha}$ is injective, $\overline{\alpha}(S)\neq\overline{\alpha}(T)$, so that $(\overline{\alpha}(S))(x)\neq(\overline{\alpha}(T))(x)$ for some $x\in C^Q$. Let the prescription $\mathfrak{N}\in\overline{K}^Q$ be such that $G(\mathfrak{N})=Q$, that is $\mathfrak{N}(q)=\odot$ for every $q\in Q$. Then $\mathfrak{N}\Delta x=x$ and $S[\mathfrak{N}]=S$. Now $x\in C^Q$, so that $x\in A^Q$ and $(\beta(S))(x)=(\beta(S))(\mathfrak{N}\Delta x)=(\overline{\alpha}(S[\mathfrak{N}]))(x)=(\overline{\alpha}(S))(x)$. Similarly $(\beta(T))(x)=(\overline{\alpha}(T))(x)$. Therefore $(\beta(S))(x)\neq(\beta(T))(x)$ and $\beta(S)\neq\beta(T)$. So β is injective and hence an embedding. Therefore the minor class $\mathcal{S}$ is embeddable in $\mathcal{F}^{\mathcal{Q}}(K,C,B)$. ☺

Theorem 8.1 was proved by constructing an embedding, and then going through the lengthy, but routine, process of verifying that it is one. A similar approach is possible for the following theorem. However a different (more elegant, I believe) proof is given using some universal algebra theory.

<u>Theorem 8.4</u>: If $(\mathcal{S},\mathcal{Q},\mathcal{J},K,\mathcal{P})$ is a minor class with $|K|\geq 2$, then there exists a set B and a congruence $\mathfrak{q}$ on $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$ such that $(\mathcal{S},\mathcal{Q},\mathcal{J},K,\mathcal{P})$ can be embedded in $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)/\mathfrak{q}$.

<u>Proof</u>: Consider the minor class $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,\{1,2\})$, abreviated to $\mathcal{F}$, and the variety $V(\{\mathcal{F}\})$ generated by it.

As shown in section 7, $V(\{\mathcal{F}\})=S(H(P(\{\mathcal{F}\})))$. And each minor class in $P(\{\mathcal{F}\})$ is isomorphic to $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$ for some set B, by the example given after the definition of direct product. Therefore, any minor class in $S(H(P(\{\mathcal{F}\})))$ is a subalgebra of a homomorphic image of $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$. Equivalently, any minor class in $S(H(P(\{\mathcal{F}\})))$ is embeddable in $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)/\mathfrak{q}$ for some set B and some congruence $\mathfrak{q}$ on $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$. (Recall that the homomorphic images of $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,B)$ are exactly the minor classes isomorphic to $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,B)/q$ for some congruence q.)

The theorem will follow if we can show that V({$\mathscr{F}$}) is the variety of ($\mathscr{Q}$,K) minor classes. But V({$\mathscr{F}$})=alg(eq({$\mathscr{F}$})). Therefore, to show that V({$\mathscr{F}$}) is indeed the variety of ($\mathscr{Q}$,K) minor classes, we need only show that eq({$\mathscr{F}$}) contains exactly the equations valid in every ($\mathscr{Q}$,K) minor class. Clearly an equation valid in every ($\mathscr{Q}$,K) minor class is valid in $\mathscr{F}$, so that the theorem can only fail if there is some equation (Y,ℓ,r) which is valid in $\mathscr{F}$, but not in every ($\mathscr{Q}$,K) minor class. Suppose that (Y,ℓ,r) is such an equation.

Now {1,2} is non-empty, so that $\mathscr{F}$ has at least one structure on every ground set, that is, $\mathscr{F}$ has full support. By the discussion on equations in section 5, the proper equation $(\mathrm{var}(\ell,r),\ell,r)$, denoted $\ell=r$, is also valid in $\mathscr{F}$, but not in every ($\mathscr{Q}$,K) minor class.

For i=1,2, let $\mathscr{F}^i$ be $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,\{i\})$. Clearly $\mathscr{F}^1$ and $\mathscr{F}^2$ are disjoint subalgebras of $\mathscr{F}$. Suppose $\ell=r$ is not a regular equation, that is, the variable x (say) on the left hand side is different from the variable y (say) on the right hand side. Choose an assignment $\alpha:\{x,y\}\to\mathscr{F}$ such that $\alpha(x)\in\mathscr{F}^1$ and $\alpha(y)\in\mathscr{F}^2$. (This is possible since, while α must be $\mathscr{Q}$ respecting, both $\mathscr{F}^1$ and $\mathscr{F}^2$ have full support, as {1} and {2} are non empty.) Now $\overline{\alpha}(\ell)\in\mathscr{F}^1$ and $\overline{\alpha}(r)\in\mathscr{F}^2$, so that $\overline{\alpha}(\ell)\neq\overline{\alpha}(r)$, meaning that $\ell=r$ is not valid in $\mathscr{F}$. (Recall from section 5 that $\overline{\alpha}$ is the extension of α.) This contradicts the original assumption, so that $\ell=r$ must be regular.

By the theory of equational deduction is universal algebra [3], there exist equations which (in the presence of the equations valid in all ($\mathscr{Q}$,K) minor classes) are valid in exactly the same ($\mathscr{Q}$,K) minor classes as $\ell=r$. These equations are obtained from $\ell=r$ by deduction rules [3] and one such equation will be $i^{\omega}p^{\mathfrak{K}}x=p^{\mathfrak{J}}x$ for some ground sets $Q,P\in\mathscr{Q}$, some prescriptions $\mathfrak{K},\mathfrak{J}\in\overline{K}^{Q}$ with $G(\mathfrak{J})=P$ and some bijection $\omega:G(\mathfrak{K})\to P$. (This equation is obtained by putting ℓ and r into canonical form, say $i^{\sigma}p^{\mathfrak{K}}x$ and $i^{\tau}p^{\mathfrak{J}}x$ respectively, and applying $i^{\tau^{-1}}$ to both sides, letting ω be $\tau^{-1}\circ\sigma$.)

The equation $i^{\omega}p^{\mathfrak{K}}x=p^{\mathfrak{J}}x$ is valid in $\mathscr{F}$, so that for every ground set $Q\in\mathscr{Q}$ and every structure $f\in\mathscr{F}_Q$ (that is, $f{:}K^Q\to\{1,2\}$), it holds that $\omega(f[\mathfrak{K}])=f[\mathfrak{J}]$. For every $f\in\mathscr{F}_Q$ and $z\in K^P$ it follows that $(\omega(f[\mathfrak{K}]))(z)=(f[\mathfrak{J}])(z)$, and by the definition of point removal and structure isomorphism in $\mathscr{F}^{\mathscr{Q}}(K,\emptyset,\{1,2\})$, that $f(\mathfrak{K}\Delta(\omega^{-1}(z)))=f(\mathfrak{J}\Delta z)$. This is only possible if for every $z\in K^P$ $(\mathfrak{K}\Delta(\omega^{-1}(z)))=(\mathfrak{J}\Delta z)$, so that $(\mathfrak{K}\Delta(\omega^{-1}(z)))(q)=(\mathfrak{J}\Delta z)(q)$ for every $q\in Q$. Now

$$(\mathfrak{K}\Delta(\omega^{-1}(z)))(q)=\begin{cases}\mathfrak{K}(q) & \text{whenever } q\in Q-G(\mathfrak{K})\\ z(\omega(q)) & \text{whenever } q\in G(\mathfrak{K}) \text{ (or equivalently, } \omega(q)\in P)\end{cases}$$

and is equal, for all $q\in Q$, to

$$(\mathfrak{J}\Delta z)(q)=\begin{cases}\mathfrak{J}(q) & \text{whenever } q\in Q-G(\mathfrak{J})=Q-P\\ z(q) & \text{whenever } q\in P.\end{cases}$$

Suppose $\omega{:}G(\mathfrak{K})\to P$ is not the identity function. Let $q\in G(\mathfrak{K})$ be such that $\omega(q)\neq q$. There are two cases to consider; either $q\in P\cap G(\mathfrak{K})$ or $q\in G(\mathfrak{K})-P$. If $q\in P\cap G(\mathfrak{K})$, then choose $z\in K^P$ by letting $z(q)$ and $z(\omega(q))$ be different elements of K (and defining z on the rest of P arbitrarily). This contradicts the above equality. If $q\in G(\mathfrak{K})-P$, then choose $z(\omega(q))$ to be an element of K, different from $\mathfrak{J}(q)$ (and define z on the rest of P arbitrarily). Again the above equality is contradicted. (Observe that both these cases use the fact that $|K|\geq 2$.) Therefore, ω is the identity, as is ω^{-1}, and $G(\mathfrak{K})=P=G(\mathfrak{J})$, so that $\mathfrak{K}(q)=\odot=\mathfrak{J}(q)$ for every $q\in P$. Also $(\mathfrak{K}\Delta z)=(\mathfrak{J}\Delta z)$ so that $\mathfrak{K}(q)=\mathfrak{J}(q)$ for every $q\in Q-P$, and hence $\mathfrak{K}=\mathfrak{J}$. But then the equation $i^{\omega}p^{\mathfrak{K}}x=p^{\mathfrak{J}}x$ is equivalent to $p^{\mathfrak{K}}x=p^{\mathfrak{K}}x$ which is valid in every $(\mathscr{Q},K)$ minor class, contradicting the assumption that (Y,ℓ,r) and hence, $\ell=r$, $i^{\omega}p^{\mathfrak{K}}x=p^{\mathfrak{J}}x$ and $p^{\mathfrak{K}}x=p^{\mathfrak{K}}x$ are not valid in every $(\mathscr{Q},K)$ minor class. So every equation valid in $\mathscr{F}$ is valid in every $(\mathscr{Q},K)$ minor class, and the result follows. ☺

SECTION 9: A CONNECTION WITH CATEGORY THEORY

There is an interesting correspondence between categories and regular unary varieties. This provides insight into the role of free algebras and their relation to the operations of the variety.

Let $\mathcal{Q}$ be a set whose elements are called *objects*. Let $\mathcal{O}=<\mathcal{O}_{s,t}\,|\,(s,t)\in\mathcal{Q}\times\mathcal{Q}>$ be a $\mathcal{Q}\times\mathcal{Q}$-partition of a set whose elements are called *morphisms*. The elements of $\mathcal{O}_{s,t}$ are *morphism from* s *to* t. And for all objects $r,s,t\in\mathcal{Q}$, let $\circ$ be a binary operation which sends each pair of morphisms $v\in\mathcal{O}_{r,s}$ and $w\in\mathcal{O}_{s,t}$, to a morphism in $\mathcal{O}_{r,t}$ denoted $w\circ v$, called the *composition* of v and w. The triple $(\mathcal{Q},\mathcal{O},\circ)$ is a *category* [4] if the following two conditions hold.

(Cat 1) For all objects $q,r,s,t\in\mathcal{Q}$, and all morphisms $u\in\mathcal{O}_{q,r}$, $v\in\mathcal{O}_{r,s}$ and $w\in\mathcal{O}_{s,t}$ it holds that $w\circ(v\circ u)=(w\circ v)\circ u$

(Cat 2) For all objects $s\in\mathcal{Q}$ there exists a morphism $1^s\in\mathcal{O}_{s,s}$, the *identity morphism on* s, such that for all objects $r,t\in\mathcal{Q}$ and all morphisms $v\in\mathcal{O}_{r,s}$ and $w\in\mathcal{O}_{s,t}$ it holds that $1^s\circ v=v$ and $w\circ 1^s=w$.

Now $\Sigma=(\mathcal{Q},\mathcal{O})$ is, of course, also a unary signature. The elements of $\mathcal{Q}$ are both objects and sorts, while the elements of $\mathcal{O}$ are both morphisms and operator symbols. Let Θ be the set of Σ-equations $wvx=(w\circ v)x$ and $1^r x=x$, where x is a variable of sort $r\in\mathcal{Q}$, for all sorts (objects) $r,s,t\in\mathcal{Q}$, and all operator symbols (morphisms) $v\in\mathcal{O}_{r,s}$ and $w\in\mathcal{O}_{s,t}$. These equations are regular, so that $\Lambda=(\Sigma,\Theta)$ is a regular unary specification and $\mathrm{alg}(\Lambda)$ is a regular unary variety. (Note again that the objects of the category are the <u>sorts</u>, not the varieties, algebras in varieties, or elements in algebras.)

An algebra A in the variety can be thought of as a "realisation" of the category in the following sense. There is a set A_s for each object (sort) $s\in\mathcal{Q}$, a function $w_A:A_s\to A_t$ for each morphism (operator symbol) $w\in\mathcal{O}_{s,t}$, for all objects (sorts) $s,t\in\mathcal{Q}$ (in particular, $1^s_A:A_s\to A_s$ is the identity function) and composition of functions acts the same way as composition of morphisms (that is, for all morphisms/operator symbols $v \in \mathcal{O}_{r,s}$ and $w \in \mathcal{O}_{s,t}$, the function $(w_A \circ v_A): A_r \to A_t$ is the same as the function $(w \circ v)_A: A_r \to A_t$, as guaranteed by the equations). In fact the algebras in the variety are exactly the "realisations" of the category in this sense.

For example, consider the category whose objects are the elements of $\mathcal{Q}$, where $\mathcal{Q}$ is a set of sets, and for each pair of sets $Q, P \in \mathcal{Q}$, the morphisms from Q to P are the bijections from Q to P, and composition of morphisms is composition of bijections. Then the regular unary variety corresponding to this category is the variety of isomorphism classes whose set of ground sets is $\mathcal{Q}$.

Let $(\mathcal{Q}, \mathcal{O}, \circ)$ be a category and let $\mathcal{V} = \mathrm{alg}(\Lambda)$ be the regular unary variety obtained from this category as above. If x is a variable of sort $s \in \mathcal{Q}$, let us determine what the free algebra of $\mathcal{V}$, freely generated by x, looks like. Recall (from section 7) that this algebra is unique up to isomorphism, and one such algebra is $T^{\Sigma}(x)/eq^{x}(\mathcal{V})$, abbreviated to **F** (where Σ is the unary signature $(\mathcal{Q}, \mathcal{O})$). Abbreviate $T^{\Sigma}(x)$ to **T**. The elements of T are the Σ-terms which have the variable x, and $eq^{x}(\mathcal{V})$ is an equivalence relation on these, making the elements of F equivalence classes of elements of T. According to the abovementioned equations, any term $w^1 w^2 \ldots w^m x$ is equivalent to the term wx where the morphism $w = w^1 \circ w^2 \circ \ldots w^m$. (For m=0, the term x is equivalent to $1^s x$.) The definition of "term" (see section 5) ensures that this composition of morphisms is defined. Clearly each equivalence class contains exactly one term of the form wx, where w is a morphism from object $s \in \mathcal{Q}$ (since x is of sort s), and all morphisms from object s arise in this way. Also if wx, and hence the equivalence class containing it, $wx \cdot eq^{x}(\mathcal{V})$, are of sort $t \in \mathcal{Q}$ then w is a morphism from object s to object t. Thus, for each sort $t \in \mathcal{Q}$, the elements of F_t are exactly the equivalence classes $wx \cdot eq^{x}(\mathcal{V})$ where $w \in \mathcal{O}_{s,t}$ is a morphism from s to t. For any sort (object) $r \in \mathcal{Q}$ and any operator symbol (morphism) $u \in \mathcal{O}_{t,r}$ the function $u_{\mathbf{T}}:T_t \to T_r$ sends wx to uwx and this is equivalent, under $eq^X(\mathscr{C})$, to the term $(u \circ w)x$ which has a single operator symbol $(u \circ w)$. Therefore the function $u_{\mathbf{F}}:F_t \to F_r$ sends $wx \cdot eq^X(\mathscr{C})$ to $uwx \cdot eq^X(\mathscr{C})$ which equals $(u \circ w)x \cdot eq^X(\mathscr{C})$.

We can discard superfluous symbols by taking the isomorphic copy, denoted $\mathbf{F}^{\mathscr{C}}(1^S)$, of the above algebra $\mathbf{F}$, induced by the isomorphism from $\mathbf{F}$ to $\mathbf{F}^{\mathscr{C}}(1^S)$ which sends each $wx \cdot eq^X(\mathscr{C})$ to w. So the elements of sort $t \in \mathcal{Q}$ of $F_t^{\mathscr{C}}(1^S)$ are exactly the elements of $\mathcal{O}_{s,t}$. Note that these are also morphisms and operator symbols. This causes no problems and in fact highlights the connection between these three kinds of entities. For any sort (object) $r \in \mathcal{Q}$ and any operator symbol (morphism) $u \in \mathcal{O}_{t,r}$, it follows (from the isomorphism from $\mathbf{F}$ to $\mathbf{F}^{\mathscr{C}}(s^S)$) that the function $u_{\mathbf{F}^{\mathscr{C}}(1^S)}:F_t^{\mathscr{C}}(1^S) \to F_r^{\mathscr{C}}(1^S)$ sends each element (morphism) $w \in F_t^{\mathscr{C}}(1^S)(=\mathcal{O}_{s,t})$ to the element (morphism) $(u \circ w) \in F_r^{\mathscr{C}}(1^S)(=\mathcal{O}_{s,r})$. (That is, $u_{\mathbf{F}^{\mathscr{C}}(1^S)}$ has the effect of composing by morphism u on the left.)

It is interesting to consider homomorphisms from $\mathbf{F}^{\mathscr{C}}(1^S)$. For any algebra $A \in \mathscr{C}$, any sort $s \in \mathcal{Q}$, and any element of $a \in A_s$ define the homomorphism $\phi_a^{\mathbf{A}}$ (abbreviated to ϕ_a when $\mathbf{A}$ is known) as follows. The homomorphism $\phi:\mathbf{F}^{\mathscr{C}}(1^S) \to A$ sends $1^S \in \mathbf{F}^{\mathscr{C}}(1^S)$ to $a \in A$, and this uniquely determines ϕ_a. For any sort $t \in \mathcal{Q}$ and any element $w \in F_t^{\mathscr{C}}(1^S)(=\mathcal{O}_{s,t})$ it follows that $\phi_a(w)=\phi_a(w \circ 1^S)=\phi_a(w_{\mathbf{F}^{\mathscr{C}}(1^S)}(1^S))=w_{\mathbf{A}}(\phi_a(1^S))=w_{\mathbf{A}}(a)$; that is, ϕ_a sends w to $w_{\mathbf{A}}(a)$.

It is natural to ask if there is a construction yielding a category from a regular unary variety, which in a certain sense, is the inverse of the earlier construction. Not every regular unary variety arises directly from a category via the earlier construction, and an attempt to use operator symbols as morphisms need not work, (since the equations may not be of the right form).

Instead, the free algebras (generated by one element) provide the key, and the construction is as follows. Let the set of objects for the category be the set of sorts $\mathcal{Q}$ (say) for $\mathcal{C}$. For each sort $s\in\mathcal{Q}$ let $F^{\mathcal{C}}(1^s)$ be a free algebra of $\mathcal{C}$, freely generated by an element 1^s of sort s. It is easily arranged that the universes $F^{\mathcal{C}}(1^s)$ be disjoint for distinct objects s. For each sort (object) $t\in\mathcal{Q}$ let $\mathcal{O}_{s,t}$, the set of morphisms from s to t, be $F_t^{\mathcal{C}}(1^s)$. Finally, for all objects (sorts) $r,s,t\in\mathcal{Q}$ and all morphisms (elements of free algebras) $w\in\mathcal{O}_{s,t}(=F_t^{\mathcal{C}}(1^s))$ and $u\in\mathcal{O}_{t,r}(=F_r^{\mathcal{C}}(1^t))$, composition of morphisms is defined by letting $u\circ w$ be $\phi_u(w)$. (Recall that the homomorphism $\phi_u:F^{\mathcal{C}}(1^r)\to F^{\mathcal{C}}(1^t)$ sends $1^r\in F_r^{\mathcal{C}}(1^r)$ to $u\in F_r^{\mathcal{C}}(1^t)$.) Since the algebras $F^{\mathcal{C}}(1^s)$ for $s\in\mathcal{Q}$ are unique up to isomorphism, the category $(\mathcal{Q},\mathcal{O},\circ)$ is uniquely determined except for the names of the morphisms; that is, it is unique up to a category isomorphism which fixes objects. Two categories which are isomorphic, with the isomorphism fixing objects, are *object–equal*. In particular, if $\mathcal{C}$ were constructed from a category as per the original construction, then the reverse construction yields a category which is object–equal to the first, (and can be chosen to be equal). Two regular unary varieties are *equivalent* if the above construction can yield the same category for both (in which case it must yield categories which are object–equal). (Variety equivalence for 1–sorted algebras is defined in [5].) So, via these constructions, there is a one–to–one correspondence between regular unary varieties (up to equivalence) and categories (up to object–equality).

When two varieties are equivalent, there is a (usually obvious) one–to–one correspondence between their algebras. Corresponding algebras have the same universe, and homomorphisms between corresponding pairs of algebras, as functions between their universes, are exactly the same for the two equivalent varieties. So most theory holding for one variety, immediately follows for the other. Nevertheless, the algebras in one variety may have a totally different signature to those in the other variety. Theory which is intrinsically dependent on the signature (for example, algorithms for term manipulation and equational deduction [3]) does not automatically transfer from one variety to an equivalent variety, but this paper is not concerned with such theory.

If $\mathcal{Q}$ is a hereditary set of finite sets and K is a set, then the variety of ($\mathcal{Q}$,K) minor classes, is equivalent to the variety of ($\mathcal{Q}$,K) point removal classes (as it is routine to show). As was observed in section 3, a ($\mathcal{Q}$,K) minor class is immediately recognisable as a ($\mathcal{Q}$,K) point removal class, and visa versa, despite the fact that these algebras have quite different signatures.

Let $\mathscr{C}$ be the (regular unary) variety of ($\mathcal{Q}$,K) minor classes for some hereditary set of (not necessarily finite) sets and some set K. A category ($\mathcal{Q}$,$\mathcal{O}$,$\circ$), unique up to object–equality can be constructed from $\mathscr{C}$, as above, where the objects are the ground sets (as well as the sorts). The morphisms (in $\mathcal{O}$) are in one–to–one correspondence with all the possible minor class operations, and these are any sequence of point removals and isomorphisms, or equivalently a point removal followed by an isomorphism (see section 5). For all objects (ground sets) $Q,P \in \mathcal{Q}$, it follows that the morphisms from Q to P are in one–to–one correspondence with the pairs $(\omega,\mathfrak{K})$ for each prescription $\mathfrak{K} \in \overline{K}^{Q}$ and each bijection $\omega : G(\mathfrak{K}) \to P$. Assume, without loss of generality, that K is disjoint from every ground set, and let $1:Q \cup K \to Q \cup K$ be the identity function. Let the morphism corresponding to $(\omega,\mathfrak{K})$ be the function $(1 \Delta \mathfrak{K} \Delta \omega):Q \cup K \to P \cup K$ and let composition of morphisms be composition of functions. It is routine to check that this gives the appropriate category. It is informative to examine the above function. Each $q \in Q - G(\mathfrak{K})$ is sent to $\mathfrak{K}(q)$, signifying that point is removed in manner $\mathfrak{K}(q)$, each $q \in G(\mathfrak{K})$ is sent to $\omega(q)$, point q is "renamed" as point $\omega(q)$, and each manner $\ell \in K$ is sent to itself, since once a point is removed (in some manner) it stays removed (in that manner). Observe that this function is surjective.

It is interesting to look at the case where all the morphisms of a category are empimorphisms. If r and s are objects, then the morphism v from r to s is an *epimorphism* if for all objects t and morphisms u and w from s to t it holds that $u \circ v = w \circ v$ implies $u=w$. Let $(\mathcal{Q}, \mathcal{O}, \circ)$ be a category in which every morphism is an epimorphism, let $\mathscr{C}$ be the regular unary variety constructed from it and for each sort (object) $r \in \mathcal{Q}$, let $F^{\mathscr{C}}(1^r)$ be the free algebra of $\mathscr{C}$, freely generated by 1^r, whose elements are morphism from r, as defined earlier. Then $\mathscr{C}$ is a *special unary variety.* For any sort (object) $s \in \mathcal{Q}$ and any element (morphism) $v \in F_s^{\mathscr{C}}(1^r)(=\mathcal{O}_{r,s})$ the homomorphism $\phi_v : F^{\mathscr{C}}(1^s) \to F^{\mathscr{C}}(1^r)$ is an embedding, since if $\phi_v(u)=\phi_v(w)$ then (by definition) $u \circ v = w \circ v$ so that $u=w$ (since v is an epimorphism). Therefore $F^{\mathscr{C}}(1^r)\mathrm{inc}(v)$ is a free algebra of $\mathscr{C}$, freely generated by v, since it is isomorphic to $F^{\mathscr{C}}(1^s)$. These implications also hold in the reverse direction making these statements equivalent. Therefore the regular unary variety $\mathscr{C}$ is a special unary variety if and only if, whenever F(a) is a free algebra of $\mathscr{C}$, freely generated by a, and $b \in F(a)$, then $F(a)\mathrm{inc}(b)$ is a free algebra of $\mathscr{C}$, freely generated by b, and this occurs exactly when ϕ_b is an embedding.

Returning to our example, the variety of $(\mathcal{Q}, K)$ minor classes is a special unary variety, since every morphism in the corresponding category is an epimorphism. (This follows immediately from the fact that the abovementioned functions $(1 \Delta \mathfrak{K} \Delta x)$, which serve as the morphisms, are surjective.)

SECTION 10: THE CORE AND COMPLETION OF A UNARY ALGEBRA

Let $\mathscr{C}$ be a special unary variety. Without loss of generality we assume that $\mathscr{C}$ is derived from a category $(\mathscr{Q},\mathscr{O},\circ)$, where $\Sigma=(\mathscr{Q},\mathscr{O})$ and $\Lambda=(\Sigma,\Theta)$, as in the previous section. While some of the material in this section works in a much more general setting, the section as a whole requires that $\mathscr{C}$ is a special unary variety. This theory is put into context for minor classes in section 11.

For each object/sort $s\in\mathscr{Q}$, recall that $\mathbf{F}^{\mathscr{C}}(1^s)$ (abbreviated to $\mathbf{F}(1^s)$, since $\mathscr{C}$ is fixed throughout this section) is the free algebra of $\mathscr{C}$, freely generated by 1^s, whose elements are the morphisms from object s, as defined in the previous section. Note that the elements of $\mathscr{O}$ are simultaneously morphisms, operator symbols, and elements of the algebras $\mathrm{F}(1^s)$ where $s\in\mathscr{Q}$.

For any algebra $\mathbf{A}\in\mathscr{C}$, any sort $s\in\mathscr{Q}$, and any element $a\in A_s$, recall that $\phi_a^{\mathbf{A}}$, or ϕ_a for short, is the homomorphism $\phi_a:\mathrm{F}(1^s)\to\mathrm{B}$ which sends each $w\in\mathrm{F}_t(1^s)(=\mathscr{O}_{s,t})$ to $w_{\mathbf{A}}(a)$. In particular, if $r\in\mathscr{Q}$ and $v\in\mathrm{F}_s(1^r)$ $(=\mathscr{O}_{r,s})$, then $\phi_v:\mathrm{F}(1^s)\to\mathrm{F}(1^r)$ sends $w\in\mathrm{F}_t(1^s)$ $(=\mathscr{O}_{s,t})$ to $w_{\mathrm{F}(1^r)}(v)$ which is $w\circ v$. Also $\phi_a\circ\phi_w$ is $\phi_{w_{\mathbf{A}}(a)}$ since $\phi_w:\mathrm{F}(1^t)\to\mathrm{F}(1^s)$ sends 1^t to w and $\phi_a:\mathrm{F}(1^s)\to\mathbf{A}$ sends w to $w_{\mathbf{A}}(a)$. In particular $\phi_v\circ\phi_w=\phi_{w_{\mathrm{F}(1^r)}(v)}=\phi_{w\circ v}$. (Note the order of v and w in the first and third expression.) For every $v\in\mathscr{O}$, ϕ_v is an embedding, since this is equivalent to the condition that $\mathscr{C}$ is a special unary variety, (see section 9).

Let $(\mathbf{A}^i\,|\,i\in I)$ be a family of algebras in $\mathscr{C}$. For each sort $s\in\mathscr{Q}$ let $A_s=\bigcup_{i\in I} A_s^i$ and suppose that $A_s\cap A_t=\phi$ for all distinct $s,t\in\mathscr{Q}$. This ensures that $A=\langle A_s\,|\,s\in\mathscr{Q}\rangle=\bigcup_{i\in I}\langle A_s^i\,|\,s\in\mathscr{Q}\rangle$ is well defined. Let us attempt to define an algebra A, with universe A, such that $\mathbf{A}^i=\mathrm{A}\mathrm{inc}\mathrm{A}^i$ for each $i\in I$. This requires that $w_{\mathbf{A}}(a)=w_{\mathbf{A}^i}(a)$ for all sorts $s,t\in\mathscr{Q}$, every operator symbol $w\in\mathscr{O}_{s,t}$, every element $a\in A_s$, and every $i\in I$ such that $a\in A_s^i$. Therefore the algebra A exists exactly when $w_{A^i}(a)$ is independent of i. In this case, the algebras $(A^i \mid i \in I)$ are *compatible*, and A is denoted by $\bigcup_{i \in I} A^i$. By definition, A is a Σ-algebra. In fact A is in $\mathscr{C}$ as the following argument shows. Suppose A is not in $\mathscr{C}$. Then some (regular unary) equation, $\ell=r$, in Θ is not valid in A. So there is some assignment sending the (single) variable in $\ell=r$ to (say) $a \in A$ such that its extension sends ℓ and r to distinct elements of A. But $a \in A^i$ for some $i \in I$, making $\ell=r$ invalid in A^i which contradicts $A^i \in \mathscr{C}$. Therefore A is indeed in $\mathscr{C}$.

Let $(A^i \mid i \in I)$ and $A = \bigcup_{i \in I} A^i$ be as above. Let $B \in \mathscr{C}$ and let $\alpha^i : A^i \to B$ be homomorphisms. Let us attempt to define a homomorphism $\alpha : A \to B$ such that $\alpha(a) = \alpha^i(a)$ for every $a \in A$ and every $i \in I$ such that $a \in A^i$. Clearly α is well defined exactly when $\alpha^i(a)$ is independent of i, for every $a \in A$. In this case, the homomorphisms $(\alpha^i \mid i \in I)$ are *compatible* and α is denoted $\bigcup_{i \in I} \alpha^i$. The fact that α is a homomorphism is easily deduced from the fact that each α^i is.

Consider algebras $A,B,C,D \in \mathscr{C}$, where $C \leq A$ and $D \leq B$, and a homomorphism $\alpha : A \to B$. The *restriction* of α to C, denoted $\alpha|_C : C \to B$, satisfies $(\alpha|_C)(c) = \alpha(c)$ for every $c \in C$. The *inverse image of* D *under* α, denoted $\alpha^{-1}(D)$ is the subuniverse of A for which $a \in \alpha^{-1}(D)$ exactly when $\alpha(a) \in D$, and $\alpha^{-1}(D)$ is the corresponding subalgebra of A. The homomorphism α *confined to* D, denoted $\alpha|^D : \alpha^{-1}(D) \to D$, satisfies $(\alpha|^D)(a) = \alpha(a)$ for every $a \in \alpha^{-1}(D)$. For any algebras $E,F \in \mathscr{C}$ and any homomorphism $\beta : E \to F$, define $\alpha = \beta$ to mean that A=E and $\alpha(a) = \beta(a)$ for every $a \in A$. (Note that B and F are allowed to be different, although they are both extensions of the homomorphic images $\alpha(A)$ and $\beta(E)$, which are identical.) For example $\alpha|^D = \alpha|_{\alpha^{-1}(D)}$. For any homomorphism γ let dom(γ) denote the *domain of* γ. (For the above α and β, dom(α)=A and dom(β)=E.) Define a relation $\leq$ on homomorphisms, as follows. If α and β are homomorphisms, then $\beta\leq\alpha$ exactly when $\mathrm{dom}(\beta)\leq\mathrm{dom}(\alpha)$ and $\beta=\alpha|_{\mathrm{dom}(\beta)}$. (For example, with homomorphisms $\alpha:A\to B$ and $\alpha^i:A^i\to B$ as in the previous paragraph, $\mathrm{dom}(\alpha^i)=A^i, \alpha^i=\alpha|_{A^i}$ and $\alpha^i\leq\alpha$ for every $i\in I$.) Clearly $\leq$ is a partial order on homomorphisms. (Note that it is unrelated to the partial order on congruences.)

For any algebras $\mathbf{A},\mathbf{B}\in\mathscr{C}$ let $h(\mathbf{A},\mathbf{B})$ be the partially ordered set of all homomorphisms $\alpha:C\to B$, where $\mathbf{C}\leq\mathbf{A}$, with the partial order on homomorphisms as above. We now show that every element of $h(\mathbf{A},\mathbf{B})$ is bounded above by a maximal element. Let $(\alpha^i|i\in I)$ be a chain in $h(\mathbf{A},\mathbf{B})$. Clearly the homomorphisms $(\alpha^i|i\in I)$ are compatible, so that $\alpha=\bigcup_{i\in I}\alpha^i$ exists, is in $h(\mathbf{A},\mathbf{B})$, and is an upper bound for the chain. The result then follows by Zorn's Lemma. Note that if $\alpha\in h(\mathbf{A},\mathbf{B})$ and $\mathrm{dom}(\alpha)=A$ then α is maximal in $h(\mathbf{A},\mathbf{B})$ but the converse need not hold.

The following lemma is essential for the development of this section.

<u>Lemma 10.1</u>: (i) If $\alpha:C\to B$ is maximal in $h(\mathbf{A},\mathbf{B})$ (so that $\mathbf{C}\leq\mathbf{A}$) and $\mathbf{D}\leq\mathbf{A}$, then $\alpha|_{C\cap D}:C\cap D\to B$ is maximal in $h(\mathbf{D},\mathbf{B})$.

(ii) If $\beta:E\to B$ is maximal in $h(\mathbf{D},\mathbf{B})$ (so that $\mathbf{E}\leq\mathbf{D}$) and $\gamma:F\to D$ is an isomorphism, then $\beta\circ\gamma$ is maximal in $h(\mathbf{F},\mathbf{B})$. (Note that $\beta\circ\gamma$ is an unambiguous abbreviation of $\beta\circ\gamma|^{\mathrm{dom}(\beta)}$.)

(iii) If $\alpha:C\to B$ is maximal in $h(\mathbf{A},\mathbf{B})$ and $\delta:F\to A$ is an embedding, then $\alpha\circ\delta$ is maximal in $h(\mathbf{F},\mathbf{B})$.

<u>Proof</u>: (i) Suppose $\beta:E\to B$ satisfies $(\alpha|_{C\cap D})\leq\beta$ and $\beta\in h(\mathbf{D},\mathbf{B})$ (so that $(\mathbf{C}\cap\mathbf{D})\leq\mathbf{E}\leq\mathbf{D}$). Then β is compatible with α, since they agree on $C\cap E=C\cap D$, so that $\alpha\cup\beta:C\cup E\to B$ exists, $\alpha\leq(\alpha\cup\beta)$ and $(\alpha\cup\beta)\in h(\mathbf{A},\mathbf{B})$. Since α is maximal in $h(\mathbf{A},\mathbf{B})$ it follows that $C\cup E=C$ and hence $E=C\cap D$, so that $\beta=(\alpha|_{C\cap D})$.

Therefore $\alpha|_{C\cap D}$ is maximal in $h(\mathbf{D},\mathbf{B})$.

(ii) For any $\beta,\varepsilon\in h(\mathbf{D},\mathbf{B})$ it is clear that $\beta\leq\varepsilon$ exactly when $\beta\circ\gamma\leq\varepsilon\circ\gamma$ ($\beta\circ\gamma$ and $\varepsilon\circ\gamma$ are in $h(\mathbf{F},\mathbf{B})$), so if β is maximal in $h(\mathbf{D},\mathbf{B})$ then $\beta\circ\gamma$ is maximal in $h(\mathbf{F},\mathbf{B})$.

(iii) Let $D=\delta(F)$, let the isomorphism $\gamma:F\to D$ be $\delta:F\to\delta(F)$, let $\mathbf{E}=C\cap\mathbf{D}$, and let $\beta=\alpha|E$. The result then follows from (i) to (ii). ☺

An algebra $\mathbf{Z}\in\mathscr{C}$ is *complete* if for every sort $s\in\mathscr{Q}$ it holds that, if the homomorphism β is maximal in $h(\mathbf{F}(1^s),\mathbf{Z})$, then $\mathrm{dom}(\beta)=\mathbf{F}(1^s)$. Observe that such a β must be ϕ_b, where $b=\beta(1^s)$. So an equivalent condition for Z to be complete is that, for any sort $s\in\mathscr{Q}$ and any homomorphism $\alpha\in h(\mathbf{F}(1^s),\mathbf{Z})$ there exists $b\in Z_s$ such that $\alpha\leq\phi_b$. In this case, for every $w\in\mathrm{dom}(\alpha)\leq F(1^s)$ it holds that $\alpha(w)=\phi_b(w)=w_{\mathbf{Z}}(b)$. (In a sense we are "solving for b".) For an algebra $\mathbf{A}\in\mathscr{C}$ it is of interest to find an extension of $\mathbf{A}$ which is complete, and to find a "smallest" such complete extension. The theory which leads to this result is also of interest.

Consider algebras $\mathbf{A},\mathbf{B}\in\mathscr{C}$ with $\mathbf{A}\leq\mathbf{B}$. It is informative to examine homomorphisms of the form $\phi_b|^A$ (or $\phi_b^{\mathbf{B}}|^A$) where $b\in B$. This is because, for $\mathbf{B}$ to be complete, it is necessary that every homomorphism α which is maximal in $h(\mathbf{F}(1^s),\mathbf{A})(\subseteq h(\mathbf{F}(1^s),\mathbf{B}))$ is bounded above by some ϕ_b, so that $\alpha=\phi_b|^A$. It turns out that there is a "smallest" complete $\mathbf{B}\geq\mathbf{A}$, with exactly one such b corresponding to each such α.) This motivates the following definition. If $b\in B_s$ (for some sort $s\in\mathscr{Q}$) and $\phi_b|^A$ is maximal in $h(\mathbf{F}(1^s),\mathbf{A})$, then b is *upon* $\mathbf{A}$. In particular, every element of A is upon $\mathbf{A}$.

For any element $b\in B_s$, and any operator symbol $w\in\mathscr{O}_{s,t}=F_t(1^s)$, it holds that $\phi_{w_{\mathbf{B}}(b)}|^A=(\phi_b\circ\phi_w)|^A=(\phi_b|^A)\circ\phi_w$. Now ϕ_w is an embedding, so by lemma 10.1, if $\phi_b|^A$ is maximal in $h(\mathbf{F}(1^s),\mathbf{A})$ then $\phi_{w_{\mathbf{B}}(b)}|^A$ is maximal in $h(\mathbf{F}(1^t),\mathbf{A})$. That is, if b is upon $\mathbf{A}$, then $w_{\mathbf{B}}(b)$ is upon $\mathbf{A}$. Therefore,

$Up(A,B)$, the set of elements of B which are upon **A**, is a subuniverse of B. Let $\mathbf{Up(A,B)}$ be the corresponding subalgebra of **B**. Clearly, if $\mathbf{A \le C \le B}$, then $\mathbf{Up(A,C) = Up(A,B) \cap C}$, and in particular $\mathbf{Up(A, Up(A,B)) = Up(A,B)}$. If $\mathbf{Up(A,B) = B}$ (that is, every $b \in B$ is upon **A**), then **B** is a *vertical extension* of **A**, and **A** is a *vertical subalgebra* of **B**, denoted $\mathbf{A \prec B}$.

<u>Theorem 10.2</u>: (i) If $\mathbf{A \le B}$ then the following are equivalent.

(a) $\mathbf{A \prec B}$.

(b) If $b \in B_s$ (for some sort $s \in \mathcal{S}$) then $\phi_b|^{\mathbf{A}}$ is maximal in $h(\mathbf{F}(1^s), \mathbf{A})$.

(c) If $\mathbf{C} \in \mathscr{C}$ and $\alpha: \mathbf{C} \to \mathbf{B}$ is a homomorphism, then $\alpha|^{\mathbf{A}}$ is maximal in $h(\mathbf{C}, \mathbf{A})$.

(d) If $b \in B_s - A_s$ (for some sort $s \in \mathcal{S}$), then there does not exist $a \in A_s$ such that $(\phi_b|^{\mathbf{A}}) \le \phi_a$.

(e) Let $\mathbf{C} \in \mathscr{C}$ and $\alpha \in h(\mathbf{C}, \mathbf{B})$. Then if $\beta \in h(\mathbf{C}, \mathbf{A})$ with $(\alpha|^{\mathbf{A}}) \le \beta$, then α and β are compatible (so that $\alpha \cup \beta$ exists).

(ii) The relation $\prec$ is a partial order.

(iii) If $\mathbf{A, B, C} \in \mathscr{C}$ with $\mathbf{A \le B \le C}$ and $\mathbf{A \prec C}$, then $\mathbf{A \prec B}$.

(iv) If $\mathbf{A, B} \in \mathscr{C}$ and $\mathbf{A \le B}$ is complete, then $\mathbf{Up(A,B)}$ is complete.

<u>Proof</u>: (i) It is immediate that (a) and (b) are equivalent, and clearly (c) implies (b). If (b) holds then (d) must hold, since otherwise $\mathrm{dom}(\phi_b|^{\mathbf{A}})$ is a proper subset of $F(1^s)$ while $\mathrm{dom}(\phi_a) = F(1^s)$, contradicting the maximality of $\phi_b|^{\mathbf{A}}$ in $h(\mathbf{F}(1^s), \mathbf{A})$.

Suppose (e) does not hold. Then there exist incompatible α and β as in (e), so that there exists $c \in C_s$ (for some sort $s \in \mathcal{S}$) such that $\alpha(c) \ne \beta(c)$. Now $c \notin \alpha^{-1}(A)$, since α and β agree there, so $\alpha(c) \in B - A$ whereas $\beta(c) \in A$. Let $b = \alpha(c)$ and $a = \beta(c)$, so that $\phi_b|^{\mathbf{A}} = (\alpha \circ \phi_c)|^{\mathbf{A}} = (\alpha|^{\mathbf{A}}) \circ \phi_c$, and $\phi_a = \beta \circ \phi_c$. Since $\alpha|^{\mathbf{A}} \le \beta$ it follows that $\phi_b|^{\mathbf{A}} \le \phi_a$, contradicting (d). Therefore (d) implies (e).

Finally suppose that (e) holds. Let α:C→B be a homomorphism, so in particular $\alpha \in$h(C,B). Let $\beta \in$h(C,A) with $\alpha|^{A} \leq \beta$. Now β only takes values in A, while outside α^{-1}(A),α only takes values in B–A. But by (e), α and β are compatible so that β must be equal to $\alpha|^{A}$. Therefore $\alpha|^{A}$ is maximal in h(C,A) and (c) holds. It follows that (a),(b),(c),(d) and (e) are equivalent.

(ii) Clearly **A**–<**B** and **B**–<**A** if and only if **A**=**B**. Suppose **A**–<**B** and **B**–<**D** (so in particular **A**≤**B**≤**D**). Consider condition (e), and let **C**∈$\mathscr{C}$, $\alpha \in$h(**C**,**D**) and $\beta \in$h(**C**,**A**) with $(\alpha|^{A}) \leq \beta$. Now $(\alpha|^{B}) \in$h(**C**,**B**) and $(\alpha|^{B})|^{A} = \alpha|^{A} \leq \beta$, so that $\alpha|^{B}$ and β are compatible, since **A**–<**B**. Therefore $(\alpha|^{B}) \cup \beta$ exists, $(\alpha|^{B}) \cup \beta \in$h(**C**,**B**) and $\alpha|^{B} \leq \alpha|^{B} \cup \beta$. Since **B**–<**D**, it follows that α and $(\alpha|^{B}) \cup \beta$ are compatible and $\alpha \cup ((\alpha|^{\beta}) \cup \beta) = \alpha \cup \beta$ exists. But then α and β are compatible so that **A**–<**D**.

(iii) If **A**≤**B**≤**C** and **A**–<**C** then **Up**(**A**,**B**)=**Up**(**A**,**C**)∩**B**=**C**∩**B**=**B**, so that **A**–<**B**.

(iv) Suppose **A**≤**B** and **B** is complete, and let **U**=**Up**(**A**,**B**). For any sort s∈$\mathscr{Q}$, consider any homomorphism $\alpha \in$h(**F**(1^{s}),**U**). Let the homomorphism β be maximal in h(F(1^{s}),A) with $\alpha|^{A} \leq \beta$. From (e) it follows that α and β are compatible so that $\alpha \cup \beta$ exists. Since **B** is complete, there exists b∈b_s such that $(\alpha \cup \beta) \leq \phi_b$, and in particular $\beta \leq \phi_b$ and $\alpha \leq \phi_b$. But since β is maximal in h(F(1^{s}),**A**) it follows that b is upon **A**, and hence b∈U. It follows that, for any sort s∈$\mathscr{Q}$ and any homomorphism $\alpha \in$h(**F**(1^{s}),**U**), there exists b∈U such that $\alpha \leq \phi_b$. Therefore **U**=**Up**(**A**,**B**) is complete. ☺

Part (iv) says that if **B** is complete and is an extension of **A**, then a "smaller" complete extension of **A**, namely **Up**(**A**,**B**) is obtained by "discarding" those elements of B which are not upon **A**. Note that the conditions of part (iii) do not imply **B**–<**C**. For example let **B**,**C**∈$\mathscr{C}$ be such that **B**≤**C** holds but **B**–<**C** does not. If **A**∈$\mathscr{C}$ is the empty algebra, then **A**≤**B**, and by (d) (there does not exist a∈A) it follows that **A**–<**B** and **A**–<**C**.

If $\mathbf{A} \neg\!< \mathbf{B}$, then define a relation $q(\mathbf{A},\mathbf{B})$ (abbreviated to q when $\mathbf{A}$ and $\mathbf{B}$ are known) on $\mathbf{B}$ as follows. For every sort $s \in \mathcal{S}$ and all elements $a,b \in B_s$, let aqb exactly when $\phi_a|^A = \phi_b|^A$. Consider any operator symbol $w \in \mathcal{O}_{s,t}(=F_t(1^s))$ and recall that $\phi_{w_{\mathbf{B}}(b)}|^A = (\phi_b|^A) \circ \phi_w$ (and similarly with b changed to a). By the definition of q, aqb implies $(w_{\mathbf{B}}(a))q(q_{\mathbf{B}}(b))$, making q a congruence. A congruence τ on $\mathbf{B}$ *preserves* $\mathbf{A}$ if $a\tau b$ and $a \in A$ implies $a=b$. Now q preserves $\mathbf{A}$, since if aqb and $a \in A$, then $\phi_a = \phi_a|^A = \phi_b|^A$ which must be ϕ_b, so that $a=b$. In fact q is the unique maximal congruence which preserves $\mathbf{A}$, as shown below. Consider any congruence τ on $\mathbf{B}$ satisfying $\tau \nleq q$, so that there exist elements $a,b \in B_s$ (for some sort $s \in \mathcal{S}$) such that $a\tau b$ while $\phi_a|^A \neq \phi_b|^A$. But then there exists $w \in F_t(1^s)(=\mathcal{O}_{s,t})$ (for some sort $t \in \mathcal{S}$) such that $\phi_a(w) = w_{\mathbf{B}}(a) \in A_t$, is different from $\phi_b(w) = w_{\mathbf{B}}(b)$, whereas $(w_{\mathbf{B}}(a))\tau(w_{\mathbf{B}}(b))$, so that τ does not preserve $\mathbf{A}$. In a sense, congruences "clump together" elements of an algebra (respecting sorts and the operations of that algebra) and the congruence q does this as much as possible while preserving $\mathbf{A}$ (by keeping its elements separate from each other, and from those in B–A). If q is the minimal congruence (that is, aqb implies $a=b$ for all $a,b \in B$) then $\mathbf{B}$ is a *small extension* of A, and $\mathbf{A}$ is a ***big subalgebra*** of $\mathbf{B}$, denoted $\mathbf{A} \prec \mathbf{B}$. (In this case, the elements of B are already "clumped together" as much as possible, subject to preserving $\mathbf{A}$.)

<u>Theorem 10.3</u>: (i) If $\mathbf{A} \prec \mathbf{B}$ then $\mathbf{A} \neg\!< \mathbf{B}$.

(ii) If $\mathbf{A} \leq \mathbf{B}$ then the following are equivalent.

(a) $\mathbf{A} \prec \mathbf{B}$.

(b) If $b \in B_s$ (for some sort $s \in \mathcal{S}$) and β is maximal in $h(\mathbf{F}(1^s),\mathbf{B})$ with $(\phi_b|^A) \leq \beta$, then $\beta = \phi_b$.

(c) If $C \in \mathscr{C}$ and $\alpha: C \to B$ is a homomorphism, and β is maximal in $h(C,B)$ with $(\alpha|^{A}) \leq \beta$, then $\beta = \alpha$. In this case, given C, $\alpha|^{A}$ *uniquely determines* α.

(d) If $b \in B_s$ (for some sort $s \in \mathscr{Q}$) and the homomorphism $\beta : F(1^s) \to B$ satisfies $(\phi_b|^{A}) \leq \beta$, then $\beta = \phi_b$.

(iii) The relation $\prec$ is a partial order.

(iv) If $A \leq B \leq C$ and $A \prec C$, then $A \prec B$ and $B \prec C$.

(v) If $B \prec C$ and $D \leq C$, then $(B \cap D) \prec D$.

(vi) If $B \prec C$ and $D \prec C$, then $(B \cap D) \prec C$.

Proof: (i) The definition of $A \prec B$ assumes $A -\!\!< B$.

(ii) Suppose that $A \prec B$ and consider (d). Let $b \in B_s$ (for some sort $s \in \mathscr{Q}$) and let $\beta: F(1^s) \to B$ satisfy $(\phi_b|^{A}) \leq \beta$. Let $a = \beta(1^s)$, so that $\beta = \phi_a$. Clearly $(\phi_b|^{A}) \leq (\beta|^{A}) = (\phi_a|^{A})$. But $\phi_b|^{A}$ is maximal in $h(F(1^s),A)$, since b is upon A, so that $(\phi_b|^{A}) = (\phi_a|^{A})$. Therefore $a q b$, so that $a = b$ (by definition of $A \prec B$) and $\beta = \phi_a = \phi_b$. Thus (a) implies (d).

Conversely, suppose that (d) holds. This clearly implies condition (d) of theorem 10.2, so that $A -\!\!< B$. If $a q b$ then $(\phi_a|^{A}) = (\phi_b|^{A})$, so that $(\phi_b|^{A}) \leq \phi_a$ and (d) implies $\phi_a = \phi_b$, so that $a = b$. Thus (d) implies (a).

Clearly (c) implies (b) and (b) implies (d). Suppose that (c) does not hold. That is, there exists a homomorphism $\alpha: C \to B$, and a homomorphism β, maximal in $h(C,B)$, with $(\alpha|^{A}) \leq \beta$ but $\alpha \neq \beta$. In particular, $\beta \nleq \alpha$ so that there exists $c \in \mathrm{dom}(\beta) (\leq C)$ such that $\alpha(c) \neq \beta(c)$. Now $\alpha|^{A} \leq \beta$ so that $(\phi_{\alpha(c)}|^{A}) = (\alpha \circ \phi_c)|^{A} = (\alpha|^{A}) \circ \phi_c \leq \beta \circ \phi_c = \phi_{\beta(c)}$, whereas $\phi_{\alpha(c)} \neq \phi_{\beta(c)}$, contradicting (d). Therefore (d) implies (c) and hence (a),(b), (c) and (d) are equivalent.

(iii) Clearly $A \prec B$ and $B \prec A$ if and only if $A = B$. If $A \prec B$ and $B \prec C$ then $A \leq C$ and for any algebra $D \in \mathscr{C}$ and any homomorphis $\alpha: D \to C$, it holds that $\alpha|^{A}$

uniquely determines $\alpha|^B$ which in turn, uniquely determines $\alpha|^C=\alpha$, so that $A \prec C$. Therefore $\prec$ is a partial order.

(iv) Suppose $A \le B \le C$ and $A \prec C$. Consider any algebra $D \in \mathscr{C}$ and any homomorphism $\alpha: D \to C$. By definition $\alpha|^A$ uniquely determine α. Clearly any homomorphism uniquely determines all its restrictions. Immediately, $\alpha|^A$ uniquely determines $\alpha|^B$, so that $A \prec B$, and $\alpha|^B$ uniquely determines α, so that $A \prec C$.

(v) Suppose $B \prec C$ and $D \le C$, and consider any algebra $E \in \mathscr{C}$ and any homomorphism $\alpha: E \to D$. Then $\alpha|^B$, which is in fact $\alpha|^{B \cap D}$, uniquely determines α. Therefore $(B \cap D) \prec D$.

(vi) Suppose $B \prec C$ and $D \prec C$. Then $D \le C$, so that $(B \cap D) \prec D$, and since $\prec$ is a partial order, $(B \cap D) \prec C$. ☺

If for any $C \in \mathscr{C}$ it were the case that the intersection of all big subalgebras of C, was again a big subalgebra, then we could conclude that there would be a unique minimal big subalgebra (namely this intersection.) However part (vi) only tells us that the intersection of two, and hence by induction, of finitely many big subalgebras, is again a big subalgebra. In fact there need not be a minimal big subalgebra. But if there is, it must be unique, by (vi).

If $A \in \mathscr{C}$ has a (unique) minimal big subalgebra, then it is called the *core* of A, and A *has a core.* (Some sufficient conditions for this to occur are given later.)

On the other hand, A always has a maximal small extension, which is unique in a certain sense, as is shown shortly.

Consider algebras $A, B \in \mathscr{C}$ with $A \multimap B$. Define a function $\Phi^{A,B}$ (abbreviated to Φ when A and B are known) sending each element of B to a homomorphism, where $\Phi^{A,B}(b)=\phi_b^B|^A$ for each $b \in B$. If $b \in B$ is of sort $s \in \mathscr{S}$, that is, $b \in b_s$, then $\phi_b|^A$ is maximal in $h(F(1^s),A)$, since b is upon A. Consider the operator symbol $w \in \mathcal{O}_{s,t}(=F_t(1^s))$, for some sort $t \in \mathscr{S}$. Recall that

$\phi_{w_B(b)}|^A=\phi_b|^A\circ\phi_w$, so that $\Phi(w_B(b))=\Phi(b)\circ\phi_w$. This suggests an algebraic structure on the set of homomorphisms which are maximal in $h(\mathbf{F}(1^s),\mathbf{A})$ for some $s\in\mathscr{S}$, for which Φ is a homomorphism.

Define an algebra $\mathbf{H}^{\mathbf{A}}$ (abbreviated to $\mathbf{H}$ when $\mathbf{A}$ is known) as follows. For each sort $s\in\mathscr{S}$, let H_s be the set of maximal homomorphisms in $h(\mathbf{F}(1^s),\mathbf{A})$. For any sort $t\in\mathscr{S}$ and any operator symbol $w\in\mathscr{O}_{s,t}(=F_t(1^s))$ let the function $w_{\mathbf{H}}:H_s\to H_t$ send each $\alpha\in H_s$ to $w_{\mathbf{H}}(\alpha)=\alpha\circ\phi_w$. Note that $\alpha\circ\phi_w$ is indeed in H_t, since α is maximal in $h(\mathbf{F}(1^s),\mathbf{A})$ and ϕ_w is an embedding so that $\alpha\circ\phi_w$ is maximal in $h(\mathbf{F}(1^t),\mathbf{A})$ by lemma 10.1. Therefore $\mathbf{H}$ is a Σ-algebra, and if all the equations in Θ are valid in $\mathbf{H}$, then $\mathbf{H}\in\mathscr{C}$. The validity of these equations follows from the fact that $\phi_{1^r}:F(1^r)\to F(1^r)$ is the identity function, and $\phi_v\circ\phi_w=\phi_{w\circ v}$ (as shown earlier) for all sorts $r,s,t\in\mathscr{S}$, and all operator symbols $v\in\mathscr{O}_{r,s}$ and $w\in\mathscr{O}_{s,t}$. Therefore $\mathbf{H}$ is indeed in $\mathscr{C}$. For any algebra $\mathbf{B}\in\mathscr{C}$ with $\mathbf{A}\!-\!\!<\mathbf{B}$, the function $\Phi^{\mathbf{A},\mathbf{B}}:\mathbf{B}\to\mathbf{H}^{\mathbf{A}}$ is a homomorphism since $\Phi(w_{\mathbf{B}}(b))=\Phi(b)\circ\phi_w=w_{\mathbf{H}}(\Phi(b))$.

The homomorphism $\Phi^{\mathbf{A},\mathbf{A}}:A\to H$ sends each $a\in A$ to $\phi_a|^A$, which equals ϕ_a, and clearly $\Phi^{\mathbf{A},\mathbf{A}}$ is an embedding. Let $\overline{A}=<\overline{A}_s|s\in\mathscr{S}>$ where $A_s\subseteq\overline{A}_s$ and $|\overline{A}_s|=|H_s|$ for each sort $s\in\mathscr{S}$. Then the injection $\Phi^{\mathbf{A},\mathbf{A}}:A\to H$ can be extended to a bijection $\eta:\overline{A}\to H$. Now η becomes an isomorphism by letting the algebra $\overline{\mathbf{A}}$ inherit the algebraic structure of $\mathbf{H}$, that is, by setting $w_{\overline{\mathbf{A}}}(a)=\eta^{-1}(w_{\mathbf{H}}(\eta(a)))$ (for all sorts $s,t\in\mathscr{S}$, every operator symbol $w\in\mathscr{O}_{s,t}$, and every element $a\in\overline{A}_s$).

Consider any $a\in\overline{A}$ of sort $s\in\mathscr{S}$ (that is, $a\in\overline{A}_s$) and let us determine what $\eta(a)=\alpha$, say, can be. The homomorphism α is maximal in $h(\mathbf{F}(1^s),\mathbf{A})$. Consider $w\in dom(\alpha)(\subseteq F(1^s))$, so that $\alpha(w)\in A$. Now $\phi_{\alpha(w)}=\alpha\circ\phi_w=w_{\mathbf{H}}(\alpha)=w_{\mathbf{H}}(\eta(a))=\eta(w_{\overline{\mathbf{A}}}(a))$. The domain of $\phi_{\alpha(w)}$, and hence of $\eta(w_{\overline{\mathbf{A}}}(a))$ is all of $F(1^s)$, so that $w_{\overline{\mathbf{A}}}(a)\in A$, since η is a bijection and only sends elements of A to such homomorphisms. But $\eta|_A$ is $\Phi^{\mathbf{A},\mathbf{A}}$, so that

$\eta(w_{\mathbf{A}}(a))=\Phi^{\mathbf{A},\mathbf{A}}(w_{\overline{\mathbf{A}}}(a))=\phi_{w_{\overline{\mathbf{A}}}(a)}$. Therefore $\alpha(w)=w_{\mathbf{A}}(a)$ which equals $\phi_a(w)$, and $\phi_a(w)\in A$. It follows that $\alpha\leq\phi_a|^A$ and hence $\eta(a)=\alpha=\phi_a|^A$ by the maximality of α in $h(\mathbf{F}(1^s),\mathbf{A})$. In particular this shows that $\mathbf{A}\text{–}<\overline{\mathbf{A}}$, so that $\Phi^{\mathbf{A},\mathbf{A}}$ is defined. Now $\Phi^{\mathbf{A},\mathbf{A}}(a)$ is also $\phi_a|^A$ so that η must be $\Phi^{\mathbf{A},\overline{\mathbf{A}}}$ (and η is uniquely determined). Since $\Phi^{\mathbf{A},\overline{\mathbf{A}}}$ is an isomorphism (as η is), it follows that for every $\alpha\in H$, there is exactly one $a\in A$ such that $\alpha=\phi_a|^A$. The algebra $\overline{\mathbf{A}}$ is the *completion* of $\mathbf{A}$. It is justified calling $\overline{\mathbf{A}}$ the, rather than a, completion, since it is unique up to an isomorphism which is the identity on A.

Consider algebras $\mathbf{A},\mathbf{B}\in\mathscr{C}$ with $\mathbf{A}\text{–}<\mathbf{B}$ and let $\overline{\mathbf{A}}$ be the completion of $\mathbf{A}$. Define the homomorphism $\Omega^{\mathbf{A},\mathbf{B}}:\mathbf{B}\to\mathbf{A}$ by $\Omega^{\mathbf{A},\mathbf{B}}=(\Phi^{\mathbf{A},\overline{\mathbf{A}}})^{-1}\circ(\Phi^{\mathbf{A},\mathbf{B}})$. Then $\Omega^{\mathbf{A},\mathbf{B}}$ sends each $b\in B$ to the unique $a\in\overline{A}$ such that $\phi_b^{\mathbf{B}}|^A=\phi_a^{\overline{\mathbf{A}}}|^A$. Observe that $(\Omega^{\mathbf{A},\mathbf{B}})|_A:A\to A$ is the identity function.

Recall the congruence $\mathfrak{q}(\mathbf{A},\mathbf{B})$ on $\mathbf{B}$. It follows immediately from its definition, that $\mathfrak{q}(\mathbf{A},\mathbf{B})$ is the kernel of $\Phi^{\mathbf{A},\mathbf{B}}$, and hence the kernel of $\Omega^{\mathbf{A},\mathbf{B}}$, since $(\Phi^{\mathbf{A},\overline{\mathbf{A}}})^{-1}$ is an isomorphism. Now $\mathbf{A}\prec\mathbf{B}$ if and only if $\mathfrak{q}(\mathbf{A},\mathbf{B})$ is the minimal congruence, which occurs exactly when $\Omega^{\mathbf{A},\mathbf{B}}$ is an embedding. In particular, $\Omega^{\mathbf{A},\overline{\mathbf{A}}}:\overline{\mathbf{A}}\to\mathbf{A}$ is an embedding (the identity) so that $\mathbf{A}\prec\overline{\mathbf{A}}$. This shows that $\overline{\mathbf{A}}$ is the unique maximal small extension of $\mathbf{A}$ (unique up to an isomorphism which fixes $\mathbf{A}$).

Suppose $\mathbf{A}\text{–}<\mathbf{B}$ and $\mathbf{B}$ is complete. If $a\in\overline{A}$ is of sort $s\in\mathscr{Q}$ (so that $a\in\overline{A}_s$) then $\phi_a|^A\in h(\mathbf{F}(1^s),\mathbf{B})$ and since $\mathbf{B}$ is complete, there exists $b\in B_s$ such that $(\phi_a|^A)\leq\phi_b$. It follows that $\phi_b^{\mathbf{B}}|^A=\phi_a^{\mathbf{A}}|^A$ so that $\Omega^{\mathbf{A},\mathbf{B}}(b)=a$. Therefore if B is complete then $\Omega^{\mathbf{A},\mathbf{B}}$ is surjective. Furthermore $\overline{\mathbf{A}}$, the completion of $\mathbf{A}$, is complete. The proof of this is the same as the proof of theorem 10.2(iv) (with U changed to $\overline{\mathbf{A}}$) except that the existence of $b\in\overline{A}$ such that $\beta\leq\phi_b$ (where β is as in the earlier proof) now follows from the fact that $\overline{\mathbf{A}}$ contains all such elements b. This discussion about the relationship between **A** and $\bar{\mathbf{A}}$ is summarised in the following theorem.

Theorem 10.4: The completion $\bar{\mathbf{A}}$ of **A** is unique up to an isomorphism which fixes **A**. Furthermore

(i) **A** is the unique (up to isomorphism fixing A) maximal small extension of **A**.

(ii) $\bar{\mathbf{A}}$ is the unique (up to isomorphism fixing A) "smallest" complete extension of **A** in the following sense. If $\mathbf{B}\in\mathscr{C}$, $\mathbf{A}\leq\mathbf{B}$ and **B** is complete, then **C**=**Up**(**A**,**B**) is complete, and there is a surjective homomorphism from **C** to $\bar{\mathbf{A}}$ fixing A, namely $\Omega^{\mathbf{A},\mathbf{B}}$, so that the homomorphic image of **C** is $\bar{\mathbf{A}}$ which is complete. ☺

According to part (ii), if **B** is a complete extension of **A**, then the "smallest" complete extension is obtained by "throwing away" redundant elements of B (namely those which are not upon **A**) and "clumping together" the rest as much as possible subject to preserving **A**.

Consider algebras $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}\in\mathscr{C}$, and a homomorphism $\alpha{:}D\to B$, where $\mathbf{A}\prec\mathbf{B}$ and $\alpha^{-1}(\mathbf{A})=\mathbf{C}\leq\mathbf{D}$. Recall from theorem 10.3(ii)(c) that α is uniquely determined by $\alpha|^{A}{:}C\to A$. (That is, if $\beta\in h(\mathbf{D},\mathbf{B})$ and $\alpha|^{A}\leq\beta$, then $\beta\leq\alpha$.) Let us examine the situation in more detail. If $d\in D$ then $\phi_{\alpha(d)}=\alpha\circ\phi_d$. Therefore $\phi_{\alpha(d)}|^{A}=(\alpha\circ\phi_d)|^{A}=\alpha|^{A}\circ\phi_d=\alpha|^{A}\circ\phi_d|^{C}$ (since $\mathrm{dom}(\alpha|^{A})=C$). Hence there exists $b\in B$ such that $\phi_b|^{A}=\alpha|^{A}\circ\phi_d|^{C}$, for example $b=\alpha(d)$, and this b is unique since $\mathbf{A}\prec\mathbf{B}$. Equivalently $\alpha(\mathbf{C})=(\alpha|^{A})(\mathbf{C})-<(1\Delta\alpha|^{A})(\mathbf{D})$ where $1{:}D\to D$ is the identity function and the algebra $(1\Delta\alpha|^{A})(\mathbf{D})=\mathbf{E}$, say, with universe $E=(D-C)\cup\alpha(C)$ (it is assumed, without loss of generality, that $(D-C)\cap\alpha(C)=\emptyset$) is defined as the homomorphic image of D by postulating that the function $(1\Delta\alpha|^{A}){:}D\to E$ is a homomorphism. Clearly $\ker(1\Delta\alpha|^{A})\leq\ker(\alpha)$ so that there exists uniquely a homomorphism $\gamma{:}E\to B$ such that $\alpha=\gamma\circ(\Delta\alpha|^{A})$. Also

$\gamma(\mathbf{E})=\alpha(\mathbf{D})$, and by theorem 10.3(v), $\alpha(\mathbf{C})=\alpha(\mathbf{D})\cap\mathbf{A}\prec\alpha(\mathbf{D})$ so that γ is uniquely determined by $\gamma|^{\alpha(\mathbf{C})}$. So γ must be $\Omega^{\alpha(\mathbf{C}),\mathbf{E}}$ since they agree on $\alpha(\mathbf{C})$ (where they are both the identity). Therefore $\alpha=\Omega^{\alpha(\mathbf{C}),\mathbf{E}}\circ(1\Delta\alpha|^{\mathrm{A}})$.

Given a homomorphism β:C→A, it is natural to ask if there exists a homomorphism α:D→B such that $\alpha|^{\mathrm{A}}=\beta$. As shown above, it is necessary that $\beta(\mathbf{C})\text{–}\!<(1\Delta\beta)(\mathbf{D})=\mathbf{E}$, say, in which case the homomorphism α:D→$\overline{\mathrm{A}}$ (where $\overline{\mathbf{A}}$ is the completion of A, so that $\mathbf{B}\leq\overline{\mathbf{B}}=\overline{\mathbf{A}}$), must be $\Omega^{\beta(\mathbf{C}),\mathbf{E}}\circ(1\Delta\beta)$. Whether or not this homomorphism α sends all of D into B must be checked by some other means. (A simple method is given to determine this when the algebras are minor classes with finite ground sets).

Let us now turn to the subject of describing the core of an algebra when certain conditions guarantee its existence. For any algebra $\mathrm{A}\in\mathscr{C}$, recall the quasi order $\leq$ on its elements. If A has an infinite descending chain $\cdots<a_3<a_2<a_1$, where $a_1\in A_s$ (for some sort $s\in\mathscr{Q}$), there exist operator symbols $f^2,f^3,\cdots$ such that $a_i=f^i_{\mathbf{A}}(a_{i-1})$ for all $i=2,3,\cdots$, but there does not exist an operator symbol g^i such that $a_{i-1}=g^i_{\mathbf{A}}(a_i)$ for any $i\in\{2,3,\cdots\}$. Let $w^1=1^s$ and for $i=2,3,\cdots$ let $w^i\in F(1^s)$ be defined inductively by $w^i=f^i_{\mathbf{F}(1^s)}(w^{i-1})$ or $f^i\circ w^{i-1}$, so that $\cdots\leq w^3\leq w^2\leq w^1$. The homomorphism ϕ_{a_1}:$F(1^s)$→A sends each w^i to $w^i_{\mathbf{A}}(a_1)$ which equals a_i. There does not exist an operator symbol g^i such that $w^{i-1}=g^i_{\mathbf{F}(1^s)}(w^i)$ for any $i\in\{2,3,...\}$ since applying ϕ_{a_1} to both sides would give $a_{i-1}=g^i_{\mathbf{A}}(a_i)$, a possibility excluded above. It follows that $\cdots w^3<w^2<w^1$ and $F(1^s)$ has an infinite descending chain. Conversely, if $\mathbf{F}(1^s)$ has an infinite descending chain, then so does some algebra in $\mathscr{C}$, namely $\mathbf{F}(1^s)$. Therefore, no algebra $\mathrm{A}\in\mathscr{C}$ has an infinite descending chain if and only if no algebra $\mathbf{F}(1^s)$ for any $s\in\mathscr{Q}$ has an infinite descending chain. It turns out that these equivalent conditions are sufficient for every algebra $\mathbf{A}\in\mathscr{C}$ to have a core.

Suppose that $\mathscr{C}$ has the above property. A few definitions are needed first. For each sort $s\in\mathcal{S}$, let $E(1^s)$ be the algebra $F(1^s)\mathrm{exc}(1^s)$. Note that excluding 1^s, not only removes 1^s from the universe $F(1^s)$, but also all the invertible morphisms (called isomorphisms in the category) since these are exactly the morphisms $u\in F(1^s)$ such that $u\leq 1^s\leq u$ (that is, $u\simeq 1^s$) in the quasi order $\leq$ on $F(1^s)$. Define a $\mathcal{S}$-respecting relation ψ on A, where $a\psi b$ exactly when $\phi_a|_{E(1^s)}=\phi_b|_{E(1^s)}$ (for every sort $s\in\mathcal{S}$ and all elements $a,b\in A_s$); that is, $w_A(a)=w_A(b)$ for every $w\in F(1^s)$ except when $w\simeq 1^s$. Note that $w<1^s$ implies $w_A(a)<a$ since otherwise $w_A(a)\simeq(a)$ and there exists an (iso)morphism $v\in\mathcal{O}$, say $u=v\circ w$, such that $u_A(a)=a$ (so that u is a morphism from s to s) and $u<1^s$, making 1^s, $u, u\circ u, u\circ u\circ u, \ldots$ an infinite descending chain. Incidentally, ψ is a congruence on A, as is easily shown, but this fact is not needed here. An element $a\in A$ is a *ψ-element* if there exists $b\in A$ with $a\psi b$ and $a\neq b$ (in which case every $c\in A$ with $a\psi c$ is a ψ-element). Let $\mathbf{A}^{\psi}=\mathbf{A}\mathrm{inc}(\{a\,|\,a\in A$ and a is a ψ-element$\})$. We show that $\mathbf{A}^{\psi}$ is the core of $\mathbf{A}$.

<u>Theorem 10.5</u>: Let $\mathcal{V}$ be a special unary variety. If no algebra in $\mathcal{V}$ has an infinite descending chain, then every algebra in $\mathcal{V}$ has a core. If $\mathbf{A}\in\mathcal{V}$, then the core of $\mathbf{A}$ is $\mathbf{A}^{\psi}$.

<u>Proof</u>: Suppose $a,b\in A_s$ (for some sort $s\in\mathcal{S}$) with $a\psi b$ and $a\neq b$, and suppose that $\mathbf{B}\in\mathcal{V}$ satisfies $\mathbf{B}\prec\mathbf{A}$ and $a\notin B$. Then $(\phi_a)^{-1}(B)\leq E(1^s)$, so that $\phi_a|^B\leq\phi_a|_{E(1^s)}=\phi_b|_{E(1^s)}\leq\phi^b$. This contradicts (ii)(d) in Theorem 10.3. Therefore, every ψ-element is in every big subalgebra of $\mathbf{A}$, so that every big subalgebra contains $\mathbf{A}^{\psi}$.

It remains to show that $\mathbf{A}^{\psi}\prec\mathbf{A}$. Suppose that $\mathbf{A}^{\psi}\prec\mathbf{A}$ does not hold. Then there exist elements $a,b\in A_s$ (for some sort $s\in\mathcal{S}$) such that $\phi_b|^{A^{\psi}}\leq\phi_a$, but $a\neq b$. Let element a be minimal with this property. If $w\in E(1^s)$, then

$w<1^s$ and by an earlier argument $w_{\mathbf{A}}(a)<a$. Hence $(\phi_b|^{A^\psi})\circ\phi_w\leq\phi_a\circ\phi_w$, so that $\phi_{w_{\mathbf{A}}(b)}|^{A^\psi}\leq\phi_{w_{\mathbf{A}}(a)}$ which implies that $w_{\mathbf{A}}(b)=w_{\mathbf{A}}(a)$ by the minimality of a. But then $b\psi a$, so that $b\in A^\psi$ and hence $\phi_b|^{A^\psi}$ is ϕ_b, whose domain is the same as that of ϕ_a, namely $F(1^s)$. Hence $\phi_b=\phi_a$, and b=a contradicting the assumption that $a\neq b$. Therefore $\mathbf{A}^\psi<\mathbf{A}$ and the result follows. ☺

Let $\mathscr{V}$ be a special unary variety, none of whose algebras has an infinite descending chain. Let **A** be an algebra in $\mathscr{V}$ and let $\mathbf{A}^\psi$ and $\overline{\mathbf{A}}$ be respectively the core and completion of **A**. The big subalgebras of **A** are exactly the algebras $\mathbf{B}\in\mathscr{V}$ such that $\mathbf{A}^\psi\leq\mathbf{B}\leq\mathbf{A}$. The small extensions of **A** are exactly the algebras $\mathbf{C}\in\mathscr{V}$ (up to isomorphism fixing A) such that $\mathbf{A}\leq\mathbf{C}\leq\overline{\mathbf{A}}$. The big subalgebras of $\overline{\mathbf{A}}$ are exactly the algebras $\mathbf{D}\in\mathscr{V}$ such that $\mathbf{A}^\psi\leq\mathbf{D}\leq\overline{\mathbf{A}}$, and these are exactly the algebras (up to isomorphism fixing $\mathbf{A}^\psi$) that are small extensions of $\mathbf{A}^\psi$. Now $\overline{\mathbf{A}}$ is the completion of all such algebras **D** (up to isomorphism fixing **D**) and $\mathbf{A}^\psi$ is the core of them all (and every **D** has the same ψ-elements as $\mathbf{A}^\psi$). In particular $\overline{\mathbf{A}}$ is the completion of $\mathbf{A}^\psi$ and $\mathbf{A}^\psi$ is the core of $\overline{\mathbf{A}}$.

We now give a method for describing algebras in $\mathscr{V}$. First find the ψ-elements of **A**, thus determining its core $\mathbf{A}^\psi$. Now $\overline{\mathbf{A}}$ is the completion of **A** and $\mathbf{A}^\psi$ and for each sort $s\in\mathscr{S}$, each element $a\in A_s$ is such that $\phi_a|^{A^\psi}$ is maximal in $h(\mathbf{F}(1^s),\mathbf{A}^\psi)$. The extent to which these elements can be visualised depends on the nature of $\mathbf{F}(1^s)$ and the number of ψ-elements (or the nature of $\mathbf{A}^\psi$). Nevertheless, $\overline{\mathbf{A}}$ is uniquely determined by $\mathbf{A}^\psi$, and hence by the ψ-elements. Recall from section 5 that if an algebra **B** has no infinite descending chain then a subalgebra $\mathbf{A}\leq\mathbf{B}$ can be expressed as BexcC with C minimal. (The elements of C are obtained by choosing from the minimal elements of B–A, one of each up to $\simeq$-equivalence.) So we can specify the "excluded elements" of **A** in $\overline{\mathbf{A}}$. The ψ-elements and excluded elements of A in $\overline{\mathbf{A}}$ uniquely determine **A** up to isomorphism. It turns out that many well known minor classes have a simple description in the above terms. The following sections apply this theory to minor classes.

SECTION 11: THE ψ-DESCRIPTION OF MINOR CLASSES WITH FINITE GROUND SETS

We confine our attention to minor classes for which all ground sets are finite (because none of these have an infinite descending chain). Let $\mathcal{Q}$ be a hereditary set of finite sets and let K be a set. Let $\mathscr{C}$ be the special unary variety of $(\mathcal{Q},K)$ minor classes, which can be treated, when convenient, as the variety of $(\mathcal{Q},K)$ point removal classes (since these varieties are equivalent). The effect of point removals and structure isomorphisms on all the structures on ground set $Q\in\mathcal{Q}$ in a minor class $\mathscr{S}\in\mathscr{C}$, uniquely determine their effect on all the structures in $\mathscr{S}$ whose ground set has the same cardinality as Q (as the equations involving structure isomorphisms guarantee). Because of this, there is no loss of generality in assuming that $\mathcal{Q}$ is the set of all finite subsets of some countably infinite set.

As shown in section 9, there is a category associated with the special unary variety $\mathscr{C}$. For each sort (ground set, object) $Q\in\mathcal{Q}$, let $\mathbf{F}(1^Q)$ be the free algebra in $\mathscr{C}$, freely generated by an element 1^Q of sort Q. For each object (ground set,sort) $P\in\mathcal{Q}$, we can let the elements of $F_P(1^Q)$ be the morphisms from Q to P (see section 9). These can be denoted by pairs $(\omega,\mathfrak{K})$ for any prescription $\mathfrak{K}\in\overline{K}^Q$ and any bijection $\omega{:}G(\mathfrak{K})\to P$. For any minor class $\mathscr{S}\in\mathscr{C}$ and any structure $S\in\mathscr{S}_Q$, on ground set Q, the homomorphism $\phi_S{:}F(1^Q)\to\mathscr{S}$ sends $(\omega,\mathfrak{K})$ to $\omega(S[\mathfrak{K}])$. Let $\mathbf{E}(1^Q)$ be $\mathbf{F}(1^Q)\mathrm{exc}(1^Q)$. It was shown in section 10 that the elements removed from $F(1^Q)$ by the exclusion of 1^Q are the isomorphisms (invertible morphisms) in the category, namely pairs $(\omega,\mathfrak{K})$ where $G(\mathfrak{K})=Q$. (Such a pair $(\omega,\mathfrak{K})$ corresponds to the removal of no points followed by a structure isomorphism, or equivalently, just a structure isomorphism.)

The *order* of a structure S is the cardinality $|G(S)|$ of its ground set. If $|G(S)|=n$ then S is an *order*-n *structure* or an n*-point structure.* Sometimes the elements of ground sets are called edges or vertices, rather than points. In these cases, the word "point" in the above definition is changed accordingly.

No minor class $\mathscr{S}\in\mathscr{C}$ has an infinite descending chain so that $\mathscr{S}$ has a core $\mathscr{S}^{\psi}$. Consider structures $S,T\in\mathscr{S}$ which have the same ground set Q. Then $S\psi T$ if and only if $\phi_S|_{E(1^Q)}=\phi_T|_{E(1^Q)}$. Equivalently, $S\psi T$ if and only if $\omega(S[\mathfrak{K}[)=\omega(T[\mathfrak{K}])$ for every $(\omega,\mathfrak{K})\in E(1^Q)$. That is, performing a given non–trivial point removal, followed by a given structure isomorphism, to both S and T, yields the same result. If $S\psi T$ and $S\neq T$ then S and T are *ψ–structures.* Now $\omega(S[\mathfrak{K}])=\omega(T[\mathfrak{K}])$ if and only if $S[\mathfrak{K}]=T[\mathfrak{K}]$, so that $S\psi T$ if and only if $S[\mathfrak{K}]=T[\mathfrak{K}]$ for every prescription $\mathfrak{K}\in\overline{K}^Q$ with $G(\mathfrak{K})\neq Q$. Now if removing any single point in any manner, from both S and T yields the same structure, then removing more points cannot make them unequal. So in fact, $S\psi T$ if and only if $S[\ell,q]=T[\ell,q]$ for every manner $\ell\in K$ and every point $q\in Q$.

For any structure $S\in\mathscr{S}$, with ground set Q, let $\psi aut(S)=\{\omega|\,\omega:Q\to Q$ is a bijection and $\omega(S)\psi S\}$. Then $\psi aut(S)=\{\omega|\,\omega:Q\to Q$ is a bijection and $(\omega(S))[\mathfrak{K}]=S[\mathfrak{K}]$ for every prescription $\mathfrak{K}\in\overline{K}^Q$ with $G(\mathfrak{K})\neq Q\}$. If $\omega(S)=S$ then $\omega(S)\psi S$, so $Aut(S)\subseteq\psi aut(S)$. It may be that $Aut(S)$ is a proper subset of $\psi aut(S)$. In this case there exists $\omega\in\psi aut(S)-Aut(S)$, so that $\omega(S)\psi S$ while $\omega(S)\neq S$ (although of course, $\omega(S)\cong S$ by definition). It is also possible that $S\psi T$ while $S\not\cong T$.

Let $\mathscr{S}^-$ be the completion of $\mathscr{S}$. The *natural excluded (iso)minors* of $\mathscr{S}$ are the excluded (iso)minors of $\mathscr{S}$ in $\mathscr{S}^-$. The use of the definite article "the" requires qualification. The excluded minors are determined up to structure isomorphism and $\mathscr{S}^-$ is unique up to minor class isomorphism fixing $\mathscr{S}$. Now the latter might seem to be a problem since the natural excluded minors of $\mathscr{S}$ are elements of $\mathscr{S}^- - \mathscr{S}$, and the elements of $\mathscr{S}^- - \mathscr{S}$ can be "anything". Nevertheless, the algebraic structure of $\mathscr{S}^-$ is fixed, and this is what is important. Actually we can locate the natural excluded minors without constructing $\overline{\mathscr{S}}$ as shown below, so the "arbitrariness" of elements of $\overline{\mathscr{S}}-\mathscr{S}$ is immaterial.

The description of a minor class in terms of its ψ–structures and natural excluded minors, is called its *ψ–description.*

Consider minor classes $\mathscr{S},\mathscr{T}\in\mathscr{C}$ where $\mathscr{S}$ is a sub minor class of $\mathscr{T}$ (for example $\mathscr{T}$ could be $\mathscr{S}$ or $\overline{\mathscr{S}}$). Let the structure $S\in\mathscr{T}$, on ground set Q, be an excluded isominor of $\mathscr{S}$ in $\mathscr{T}$. The isominors of S which are in $\mathscr{S}$ are exactly those which are isomorphic to a proper minor of S. Equivalently, the elements of $F(1^Q)$ that the homomorphism $\phi_S:F(1^Q)\to\mathscr{T}$ sends into $\mathscr{S}$, are precisely the elements of $E(1^Q)$. Therefore $\phi_S{}^{-1}(\mathscr{S})$ is $E(1^Q)$ and $\phi_S|^{\mathscr{S}}:E(1^Q)\to\mathscr{S}$ is equal to $\phi_S|_{E(1^Q)}:E(1^Q)\to\mathscr{S}$.

Consider a homomorphism $\alpha:E(1^Q)\to\mathscr{S}$ (where $Q\in\mathscr{Q}$). For any minor class $\mathscr{T}\in\mathscr{C}$, where $\mathscr{S}$ is a sub minor class of $\mathscr{T}$, let $\mathscr{T}_\alpha$ be the set of all structures $T\in\mathscr{T}_Q$ for which α is equal of $\phi_T|_{E(1^Q)}$ (recall that ϕ_T is a homomorphism from $F(1^Q)$ to $\mathscr{T}$). Every structure isomorphic to a proper minor of any $T\in\mathscr{T}_Q$, is in $\alpha(E(1^Q))$ and hence in $\mathscr{S}$, so that each $T\in\mathscr{T}_Q$ is either in $\mathscr{S}$, or is an excluded isominor of $\mathscr{S}$ in $\mathscr{T}$. Also, for any structures $T,U\in\mathscr{T}_\alpha$ it holds that $\phi_T|_{E(1^Q)}=\alpha=\phi_U|_{E(1^Q)}$ so that $T\psi U$ by definition. If $\mathscr{T}_\alpha$ has more than one element then all these elements are ψ–structures of $\mathscr{T}$. Every ψ–structure $T\in\mathscr{T}$ arises in this way, simply by letting α be $\phi_T|_{E(1^Q)}$, so that $\mathscr{T}_\alpha=(T\psi)=\{U\,|\,U\in\mathscr{T}\text{ and }T\psi U\}$.

Since $\overline{\mathscr{S}}$ is complete, $\overline{\mathscr{S}}_\alpha$ is non empty for any homomorphism $\alpha:E(1^Q)\to\mathscr{S}$, where $Q\in\mathscr{Q}$ (since by definition, there exists $T\in\overline{\mathscr{S}}_Q$ such that $\alpha\leq\phi_T$, and hence $\alpha=\phi_T|_{E(1^Q)}$). In particular if T is a natural excluded isominor of $\mathscr{S}$, then $\overline{\mathscr{S}}_\alpha$ has exactly one element (since otherwise T would be a ψ–structure, so that $T\in\mathscr{S}^\psi$, contradicting the fact that $T\notin\mathscr{S}$). While $\overline{\mathscr{S}}_\alpha$ is non–empty, it is possible that $\mathscr{S}_\alpha$ is empty. In this case, the unique $T\in\overline{\mathscr{S}}_\alpha$, is not in $\mathscr{S}$, and hence is an excluded isominor of $\mathscr{S}$ in $\overline{\mathscr{S}}$, that is, T is a natural excluded minor of $\mathscr{S}$. Conversely, every natural excluded minor arises in this way, simply by letting α be $\phi_T|_{E(1^Q)}$.

Let us summarise. For all ground sets $Q \in \mathscr{Q}$, consider all homomorphisms $\alpha: E(1^Q) \to \mathscr{S}$. If $\mathscr{S}_\alpha$ has more than one element, then all its elements are ψ–structures (all ψ–equivalent to each other, and to no other structure). Every ψ–structure arises in this way. Whenever $\mathscr{S}_\alpha$ is non–empty, this indicates the existence of a natural excluded isominor, say S^α, the natural excluded isominor *indicated* by α, such that α is equal to $\phi_{S^\alpha}|_{E(1^Q)}$. (Actually S^α need not be named. It is enough to know it exists.) Every natural excluded isominor arises in this way. Natural excluded minors are obtained by partitioning the natural excluded isominors by structure isomorphism, and choosing one from each partition. Thus, ψ–structures and natural excluded (iso)minors are essentially different versions of the same concept. This method locates every instance of both of these, which is precisely the information needed for the ψ–description of $\mathscr{S}$, without constructing $\bar{\mathscr{S}}$. This also gives an independent definition of natural excluded minor, so that a minor class is complete, if and only if it has no natural excluded minors. Actually the theory outlined above works for all special unary varieties $\mathscr{C}$ in which no algebra has an infinite descending chain, as does most of the theory in this section.

For any minor class $\mathscr{T} \in \mathscr{C}$, where $\mathscr{S}$ is a sub minor class of $\mathscr{T}$, the natural excluded (iso)minors of $\mathscr{S}$ are very useful for finding the excluded minors of $\mathscr{S}$ in $\mathscr{T}$. For any natural excluded isominor of $\mathscr{S}$, which is indicated by the homomorphism $\alpha: E(1^Q) \to \mathscr{S}$, the structures in $\mathscr{T}_\alpha$ are said to be *ψ–equivalent* to this natural excluded minor. It follows from the above discussion that every natural excluded isominor of $\mathscr{S}$ in $\mathscr{T}$ is ψ–equivalent to either a structure in $\mathscr{S}$ or a natural excluded isominor of $\mathscr{S}$.

The search for the ψ-structures and natural excluded minors of $\mathscr{S}$ is made easier by the following simple observation. Any homomorphism $\alpha{:}\mathrm{E}(1^{\mathrm{Q}})\to\mathscr{S}$ is uniquely determined by specifying the effect of α, only on pairs $(\omega,\mathfrak{K})$ for which $\omega{:}\mathrm{G}(\mathfrak{K})\to\mathrm{G}(\mathfrak{K})$ is the identity function (since the set of such pairs generates $\mathrm{E}(1^{\mathrm{Q}})$). In this case $\alpha((\omega,\mathfrak{K}))$ is abbreviated to $\alpha(\mathfrak{K})$. This is equivalent to specifying the pairs $(\mathfrak{K},\alpha(\mathfrak{K}))$ for all $\mathfrak{K}\in\overline{\mathrm{K}}^{\mathrm{Q}}$ with $\mathrm{G}(\mathfrak{K})\neq\mathrm{Q}$. When K is finite, as it typically is, there are only finitely many such pairs, although this increases exponentially with the cardinality of Q. As shown earlier, all the structures $\mathrm{S}\in\mathscr{S}_{\mathrm{Q}}$ satisfying $\mathrm{S}[\mathfrak{K}]=\alpha(\mathfrak{K})$, for all $\mathfrak{K}\in\overline{\mathrm{K}}^{\mathrm{Q}}$ with $\mathrm{G}(\mathfrak{K})\neq\mathrm{Q}$, are ψ-equivalent, and if there is no such S, then α (or this restricted version of α) indicates the presence of a natural excluded isominor.

Define a $(\mathscr{Q},\mathrm{K})$ *pseudo minor class* to be a quadruple $(\mathscr{S},\mathscr{Q},\mathrm{K},\mathscr{P})$ satisfying conditions (M4) and (M5), as for minor classes. (Note that (M3) is not, and cannot, be imposed in the absence of structure isomorphisms.) By simple adaption of the discussion about $(\mathscr{Q},\mathrm{K})$ minor classes, it is clear that the class of $(\mathscr{Q},\mathrm{K})$ pseudo minor classes forms a special unary variety. The algebra with universe $\overline{\mathrm{K}}^{\mathrm{Q}}$, and with point removal defined by $\mathfrak{J}[\mathfrak{L}]=\mathfrak{J}\Delta\mathfrak{L}$ (with $\mathfrak{J}$ and $\mathfrak{L}$ as in condition (M5), section 3) is the free algebra in this variety, freely generated by an element of sort Q. A $(\mathscr{Q},\mathrm{K})$ *proper pseudo minor class* $(\mathscr{S},\mathscr{Q},\mathrm{K},\mathscr{P})$ is a pseudo minor class obtained from a minor class $(\mathscr{S},\mathscr{Q},\mathscr{J},\mathrm{K},\mathscr{P})$ by "discarding" the structure isomorphisms (which are contained in $\mathscr{J}$). The class of $(\mathscr{Q},\mathrm{K})$ proper pseudo minor classes does not form a variety since it is not closed under subalgebras (which need not be closed under structure isomorphism) nor under (pseudo minor class) homomorphic images (as the homomorphisms need not respect structure isomorphism). Nevertheless, the ψ-structures natural excluded isominors, core and completion of a proper pseudo minor class, are the same as those for the corresponding minor class. The motivation for this observation is that pseudo minor class homomorphisms are simpler to deal with than minor class homomorphisms. It must be noted, however, that while the natural excluded minors of a minor class consist of only one of each natural excluded isominor, up to isomorphism, it is necessary to exclude all of them for the corresponding proper pseudo minor class, since excluding a structure no longer removes all isomorphic structures, because there is no structure isomorphism. So if a minor class has finitely many (but not zero) natural excluded minors, the corresponding (proper) pseudo minor class has inifinitely many.

Consider a $(\mathscr{L},K)$ minor class $\mathscr{S}$ where $|K|=2$ (the most common case). A useful visualisation of the structures in $\mathscr{S}$ is obtained by modifying the "coloured grid" visualisation of section 4. Consider a structure $S \in \mathscr{S}$, with ground set Q. The Q–*hypercube* (also called a $|Q|$–hypercube) is the set of points in $\mathbb{R}^Q$ which have all their coordinates in $[0,1]=\{x \mid x \in \mathbb{R} \text{ and } 0 \leq x \leq 1\}$, that is, the Q–hypercube is $[0,1]^Q$. Assume that $K=\{0,1\}$. (Usually K is {delete,contract}, so we shall associate delete with 0 and contract with 1.) For any prescription $\mathfrak{K} \in \overline{K}^Q$, the $\mathfrak{K}$–face (also called a $|G(\mathfrak{K})|$–face) consists of all points $(x_q \mid q \in Q) \in \mathbb{R}^Q$ such that $x_q = \mathfrak{K}(q)$ whenever $q \in Q - G(\mathfrak{K})$ and $0 < x_q < 1$ whenever $q \in G(\mathfrak{K})$. So the Q–hypercube is the disjoint union, over all $\mathfrak{K} \in \overline{K}^Q$, of its $\mathfrak{K}$–faces. The $\mathfrak{K}$–subhypercube consists of all points $(x_q \mid q \in Q) \in \mathbb{R}^Q$ such that $x_q = \mathfrak{K}(q)$ whenever $q \in Q - G(\mathfrak{K})$ and $0 \leq x_q \leq 1$ whenever $q \in G(\mathfrak{K})$, and this can be identified with the $G(\mathfrak{K})$–hypercube obtained by projecting each $(x_q \mid q \in Q) \in \mathbb{R}^Q$ to $(x_q \mid q \in G(\mathfrak{K})) \in \mathbb{R}^Q$. By "drawing patterns" on each $\mathfrak{K}$–face, with the same automorphisms as $S[\mathfrak{K}]$, and with non–isomorphic patterns for non–isomorphic structures, then it will follow that these "patterned" hypercubes will faithfully represent the structures in $\mathscr{S}$, with the $\mathfrak{K}$–subhypercubes representing the appropriate minors, as in the proof of the embedding theorem 10.1. For example, if a patterned (3–dimensional) cube represents a 3–point structure, then the six (2–dimensional) squares represent the six 2–point minors obtained by removing one of three points in one of two manners, the twelve (1–dimensional) edges represent the twelve 1–point minors obtained by removing two of the three points, each in one of two manners, and the eight (0–dimensional) vertices represent the eight 0–point structures obtained by removing the three points, each in one of two manners.

The *hollow* Q–*hypercube,* is a Q–hypercube with its only $|Q|$–face (the $\mathfrak{K}$–face where $G(\mathfrak{K})=Q$, or equivalently $\mathfrak{K}\in\{\odot\}^{Q}$), called its *central face,* removed. Equivalently the hollow Q–hypercube is the disjoint union of the $\mathfrak{K}$–faces of the Q–hypercube, for $\mathfrak{K}\in(\overline{K}^{Q}-\{\odot\}^{Q})$. The pseudo minor class homomorphism $\alpha{:}\overline{K}^{Q}-\{\odot\}^{Q}\to\mathscr{S}$ can be depicted as a "patterned" hollow Q–hypercube, by drawing the pattern for $\alpha(\mathfrak{K})$ on the $\mathfrak{K}$–face, for every $\mathfrak{K}\in\overline{K}^{Q}-\{\odot\}^{Q}$.

Here is a scheme to inductively obtain the patterned Q–hypercube for each structure in $\mathscr{S}$. Suppose the patterns are known for all structures of order less than n. (This is vacuously true for n=0.) For some Q with $|Q|=n$ (all other ground sets with this cardinality are reached by structure isomorphism) consider all psuedo minor class homomorphisms $\alpha{:}\overline{K}^{Q}-\{\odot\}^{Q}\to\mathscr{S}$. For any such α, let $\mathscr{S}_{\alpha}$ be the set of all structures S such that $\phi_{S}|_{\overline{K}^{Q}-\{\odot\}^{Q}}$ equals α (where the pseudo minor class homomorphism $\alpha_{S}{:}\overline{K}^{Q}\to\mathscr{S}$ sends each prescription $\mathfrak{K}\in\overline{K}^{Q}$ to $S[\mathfrak{K}]$). For each $S\in\mathscr{S}_{\alpha}$, the pattern of the $\mathfrak{K}$–face of the corresponding Q–hypercube is the same as on the hollow Q–hypercube for α, and these are already known by induction, so it remains to draw a pattern on the central face. If there is a unique $S\in\mathscr{S}_{\alpha}$ (so that S is not a ψ–structure) then S *is uniquely determined by its proper minors*, and there is no need to draw any pattern on the central face at all. If there is more than one structure in $\mathscr{S}_{\alpha}$, then the elements of $\mathscr{S}_{\alpha}$ are ψ–equivalent to each other, and are ψ–structures. In this case, a pattern is required on the central face which is non–isomorphic for non–isomorphic structures in $\mathscr{S}_\alpha$, and which has the required automorphisms for each structure $S\in\mathscr{S}_\alpha$. (It is sufficient that the group of automorphisms of the pattern is Aut(S), but since the hollow Q–hypercube already has ψaut(S) as its group of automorphisms, it is only necessary that the intersection of the pattern's automorphisms with ψaut(S) is Aut(S).) By induction, the patterned Q–hypercube for S only has patterns on the $\mathfrak{K}$ faces for which $S[\mathfrak{K}]$ is a ψ–structure (the rest are left blank). If there are only a few ψ–structures (up to structure isomorphism) then these Q–hypercubes will be sparsely patterned.

The structures in the completion $\overline{\mathscr{S}}$, of $\mathscr{S}$, are represented by all the possible patterned hypercubes subject to two consistency conditions. (1) For any face which has a pattern corresponding to the ψ–structure S say, the patterned subhypercube which has this face as its centre, is the patterned hypercube representing S. (2) If S is a ψ–structure, on ground set Q, and if the hollow Q–hypercube corresponding to $\phi_S|_{\overline{K}}Q_{-\{\odot\}}Q$ appears, then its centre has the pattern corresponding to a structure ψ–equivalent to S. Again the situation is simple if there are not many ψ–structures.

Now if the pseudo minor class homomorphism $\alpha{:}\overline{K}^Q{-}\{\odot\}^Q{\to}\mathscr{S}$ is such that $\mathscr{S}_\alpha$ is empty, then α indicates a natural excluded isominor. The patterned hypercubes depicting structures in $\mathscr{S}$ are exactly those depicting structures in $\overline{\mathscr{S}}$, for which no patterned hollow hypercube, corresponding to any α that indicates a natural excluded minor, appears as a hollow subhypercube. Examples in the following sections clarify this situation.

Let $\mathscr{S}$ and $\mathscr{T}$ be $(\mathscr{Q},K)$ minor classes with $\mathscr{S}$ a sub minor class of $\mathscr{T}$. It is routine to show that $\mathscr{S}\,{-}{<}\,\mathscr{T}$ if and only if $\mathscr{S}$ is closed under ψ–equivalence in $\mathscr{T}$. (That is, there do not exist structures $S\in\mathscr{S}$ and $T\in\mathscr{T}-\mathscr{S}$ such that $S\psi T$.) In particular, if for some non–negative integer n, $\mathscr{S}$

contains exactly the structures of $\mathscr{T}$ with order less than n, then $\mathscr{S}$ is clearly closed under ψ-equivalence in $\mathscr{T}$, so that $\mathscr{S} -\!\!< \mathscr{T}$.

Suppose $\mathscr{T}$ and $\mathscr{S}$ are $(\mathscr{Q}, K)$ minor classes and $\alpha: \mathscr{T} \to \mathscr{S}$ is a minor class homomorphism. Recall from theorem 10.4, that α is uniquely determined by specifying the action of α on the inverse image, $\alpha^{-1}(\mathscr{S}^{\psi})$, of the core, $\mathscr{S}^{\psi}$, of $\mathscr{S}$. Conversely if $\mathscr{R}$ is a subminor class of $\mathscr{T}$ and $\beta: \mathscr{R} \to \mathscr{S}$ is given (for example $\alpha|^{\mathscr{S}^{\psi}}: \alpha^{-1}(\mathscr{S}) \to \mathscr{S}^{\psi}$) then there exists uniquely $\gamma: \mathscr{T} \to \bar{\mathscr{S}}$ such that $\gamma|^{\mathscr{S}} = \beta$ provided that $\beta(\mathscr{R}) -\!\!< (1\Delta\beta)(\mathscr{R})$ and β is maximal in $h(\mathscr{T}, \mathscr{S})$ (in particular γ is α when β is $\alpha|^{\mathscr{S}^{\psi}}$). This is guaranteed in the case that $\mathscr{R}$ consists of all the structures in $\mathscr{T}$ of order less than n, and the ψ-structures of $\mathscr{S}$ have order less than n, for some non-negative integer n. In this case $\mathscr{T}$ can be replaced by $\bar{\mathscr{T}}$, and there exists a unique homomorphism $\gamma: \mathscr{T} \to \bar{\mathscr{S}}$ such that $\gamma|^{\mathscr{S}} = \beta$. (Usually $\beta: \mathscr{R} \to \mathscr{S}^{\psi}$ is given as a homomorphism into the care of a minor class, so that $\gamma|^{\mathscr{S}^{\psi}} = \beta$ when γ exists.) To check whether or not γ sends all of $\mathscr{T}$ into $\mathscr{S}$, it suffices to check that no structure in $\mathscr{T}$ is sent by γ, to a natural excluded minor of $\mathscr{S}$ (assuming these are known, which they often are).

It is interesting to determine (or attempt to determine) the ψ-description of direct products, sub minor classes, and homomorphic images of minor classes whose ψ-description is known. For direct products this is easy. Consider a family $(\mathscr{S}^i \mid i \in I)$ of $(\mathscr{Q}, K)$ minor classes, for which the ψ-description of each $\mathscr{S}^i$ is known. Let $\mathscr{S} = \prod_{i \in I} \mathscr{S}^i$. For any family $(\alpha^i: E(1^Q) \to \mathscr{S}^i \mid i \in I)$ of homomorphisms, let the homomorphism $\alpha: E(1^Q) \to \mathscr{S}$ send each $(\omega, \mathfrak{K}) \in E(1^Q)$ to $(\alpha^i((\omega, \mathfrak{K})) \mid i \in I)$. Then $\mathscr{S}_\alpha = \prod_{i \in I} \mathscr{S}^i_{\alpha^i}$, and in particular, $|\mathscr{S}_\alpha| = \prod_{i \in I} |\mathscr{S}^i_{\alpha^i}|$. The natural excluded isominors of $\mathscr{S}$ are the structures $(S^i \mid i \in I)$ such that S^i is a natural excluded isominor of $\mathscr{S}^i$ for some $i \in I$, and S^i is either in $\mathscr{S}$ or a natural excluded isominor of $\mathscr{A}$, for every $i \in I$. (In particular, if each member of $(\mathscr{A}^i \mid i \in I)$ is complete, and so has no excluded isominors, then neither does $\mathscr{A}$, so that $\mathscr{A}$ is complete. More generally, the product of complete algebras is complete). The ψ-structures of $\mathscr{A}$ are structures $(S^i \mid i \in I)$ such that S^i is a ψ-structure for some $i \in I$, and S^i is in $\mathscr{A}$ for every $i \in I$. Two structures $(S^i \mid i \in I)$ and $(T^i \mid i \in I)$ in $\mathscr{A}$, are ψ-equivalent exactly when $S^i \psi T^i$ for every $i \in I$.

Consider $(\mathscr{Q}, K)$ minor classes $\mathscr{A}$ and $\mathscr{T}$, with $\mathscr{A}$ a sub minor class of $\mathscr{T}$. If the ψ-description of $\mathscr{T}$ is known, then the main step in finding the ψ-description of $\mathscr{A}$ is to find the excluded minors of $\mathscr{A}$ in $\mathscr{T}$, which is not necessarily easy. (It would be easy if the natural excluded minors of $\mathscr{A}$ were known, but that is one thing we wish to determine.) Suppose the excluded minors of $\mathscr{A}$ in $\mathscr{T}$ have been found. The rest is easy. First add these to the list of natural excluded minors of $\mathscr{T}$, and discard those which are not minimal. This gives the excluded minors of $\mathscr{A}$ in $\bar{\mathscr{T}}$, the completion of $\mathscr{T}$. This will not be the completion of $\mathscr{A}$ if some of the excluded minors of $\mathscr{A}$ in $\mathscr{T}$, are in the core $\mathscr{T}^{\psi}$, of $\mathscr{T}$. In this case, some ψ-structures of $\mathscr{T}$ are removed by the exclusion of these minors. If any excluded minor of $\mathscr{A}$ in $\mathscr{T}$ happens to be a ψ-structure (or possibly several isomorphic ψ-equivalent ψ-structures) of $\mathscr{T}$, then these "cancel each other out" and both are discarded. Finally, a ψ-structure in $\mathscr{T}$, which remains in $\mathscr{A}$, but for which all other structures ψ-equivalent to it are outside $\mathscr{A}$, is not a ψ-structure in $\mathscr{A}$.

The problem of finding the ψ-description of a given homomorphic image of a minor class whose ψ-description is known, is very difficult if not impossible at any reasonable level of generality. (Nevertheless it will become apparent, after reading a few more sections, that this is a problem worth pursuing.) Consider $(\mathscr{Q}, K)$ minor classes $\mathscr{T}$ and $\mathscr{A}$, and a homomorphism $\alpha: \mathscr{T} \to \mathscr{A}$, where the ψ-description of $\mathscr{T}$ is known. Suppose also that the ψ-structures of $\mathscr{A}$ are known so that α is uniquely determined by $\alpha|^{\mathscr{S}^{\psi}}$ (or by specifying which structures in $\mathscr{T}$ are sent to which ψ-structures in $\mathscr{S}$). This gives a canonical method for describing any homomorphism into $\mathscr{S}$, and puts the problem into a standard form. But while the natural excluded minors of $\alpha(\mathscr{T})$ are uniquely determined, and in principle they could be derived from the ψ-description of $\mathscr{T}$ and the homomorphism $\alpha|^{\mathscr{S}^{\psi}}$, there superficially appears to be no rhyme or reason to this situation. Examples in the following sections emphasise this. Section 15 gives some suggestions which scratch the surface of this problem.

The above three problems are each expressed in a standard form, and so in principle could have a standard solution. (This was found for two of them.) However the initial problem of finding the ψ-description of an arbitrary ($\mathscr{Q}$,K) minor class $\mathscr{S}$ is not in a standard form, since there is no limit to how $\mathscr{S}$ can be described. The only way to solve this problem is "by hook or by crook". (Most mathematicians are familiar with this method.)

The following three sections find the ψ-description of many well known minor classes and a variety of approaches is used. The ψ-structures are always found first. (This tends to be the easy part of the problem.) Sometimes they are found using the definition, and sometimes they are found by observation. Showing that a given list of ψ-structures of a minor class $\mathscr{S}$ is exhaustive, is usually achieved by demonstrating that any structure S in $\mathscr{S}$, is uniquely determined by specifying those minors of S that are ψ-structures. This amounts to showing that ϕ_S is uniquely determined by its confinement to the sub minor class of $\mathscr{S}$, generated by the (allegedly exhaustive) list of ψ-structures of $\mathscr{S}$, so that it must be the core of $\mathscr{S}$ and the list is indeed exhaustive. Natural excluded minors are trickier because they are not in $\mathscr{S}$, and there is some choice about how to describe them. Finding them uses a mixture of the two equivalent definitions (either excluded minors of $\mathscr{S}$ in $\mathscr{S}^{-}$, or the fact that they are indicated by (pseudo) minor class homomorphisms of a certain form). Demonstrating that they have all been found involves exhausting all possibilities "by hook or by crook". Often an upper bound on the order of ψ–structures and natural excluded minors is found, at which stage the problem is close to solved. Of course, any information obtained by considering related minor classes is used whenever possible.

Particularly worthy of mention is the following. Suppose that $\alpha:\mathscr{S}\rightarrow\mathscr{S}$ is a minor class automorphism of $\mathscr{S}$. Then also $\alpha|^{\mathscr{S}^{\psi}}:\mathscr{S}^{\psi}\rightarrow\mathscr{S}^{\psi}$ is an automorphism of the core $\mathscr{S}^{\psi}$, of $\mathscr{S}$, and this extends uniquely to an automorphism $\overline{\alpha}:\mathscr{S}^{-}\rightarrow\mathscr{S}^{-}$ of the completion $\mathscr{S}^{-}$, of $\mathscr{S}$. Conversely, any automorphism of $\mathscr{S}^{-}$ restricted (or equivalently, in this case, confined) to $\mathscr{S}^{\psi}$, is an automorphism of $\mathscr{S}^{\psi}$. So the automorphisms of $\mathscr{S}^{\psi}$ and $\mathscr{S}^{-}$ are in one–to–one correspondence, while in general, $\mathscr{S}$ will only have some of these. The automorphism $\overline{\alpha}$ permutes the natural excluded minors of $\mathscr{S}$, and knowledge of this symmetry can greatly reduce the search for natural excluded minors. Conversely, if $\overline{\alpha}$ is an automorphism of $\mathscr{S}^{-}$ which permutes the natural excluded minors of $\mathscr{S}$, then $\overline{\alpha}|_{\mathscr{S}}$ is an automorphism of $\mathscr{S}$. Similar comments apply to mixed automorphisms (see section 6).

Perhaps the simplest $(\mathscr{Q},\mathrm{K})$ minor class, in terms of the theory of this section is $\mathscr{F}^{\mathscr{Q}}(\mathrm{K},\emptyset,\mathrm{B})$, where B is a set (see section 4). Assume $|\mathrm{B}|\geq 2$ and $|\mathrm{K}|\geq 1$ to avoid trivial cases. The 0–point structures in $\mathscr{F}^{\mathscr{Q}}(\mathrm{K},\emptyset,\mathrm{B})$ are functions from the one element set $\mathrm{K}^{\emptyset}$ to B, and so are in one–to–one correspondence with the elements of B. The 0–point structures are all ψ–equivalent and hence are ψ–structures, since there are $|\mathrm{B}|$ of them. (In fact the 0–point structures, in any minor class, are always ψ–equivalent.) Let $f\in\mathscr{F}^{\mathscr{Q}}(\mathrm{K},\emptyset,\mathrm{B})$ be a structure on ground set Q. Now $f\in\mathscr{F}^{\mathscr{Q}}(\mathrm{K},\emptyset,\mathrm{B})$ is uniquely determined by all the values $f(y)\in\mathrm{B}$, where $y\in\mathrm{K}^{\mathrm{Q}}$, and each $f(y)$ can be chosen independently as an element of B. But since $K^Q \subseteq \overline{K}^Q$, any $y \in K^Q$ is a prescription $y \in \overline{K}^Q$ with $G(y)=\emptyset$, so that f[y] is well defined, and is the 0–point structure $f[y]:K^{\emptyset} \to B$. Let x be the unique element of $K^{\emptyset}$ so that $f(y)=f(y\Delta x)=(f[y])(x)$ for every $y \in \overline{K}^Q$ with $G(y)=\emptyset$. Therefore f is uniquely determined by all the values (f[y])(x), and hence by all 0–point minors f[y], where $y \in \overline{K}^Q$ with $G(y)=\emptyset$. That is, f is uniquely determined by its 0–point minors so that the 0–point structures are the only ψ-structures. (The order zero subgrids of the "coloured grid" for f, namely all the single gridpoint (zero dimensional) "coloured subgrids" determine the whole coloured grid.) Also each (f[y])(x) can be chosen independently as an element of B, and hence each f[y] can be chosen independently as a function from $K^{\emptyset}$ to B, for $y \in \overline{K}^Q$ with $G(y)=\emptyset$, so that there are no natural excluded minors and $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$ is complete. (The single gridpoint "coloured subgrids" can be coloured independently of each other.) This reasoning works even when $\mathcal{Q}$ contains infinite sets, so that $\mathcal{F}^{\mathcal{Q}}(K,\emptyset,B)$ is complete and still has a core (consisting of the 0–point structures) even though it also has an infinite descending chain when $|K| \geq 2$ and $|B| \geq 2$. Readers may like to show that $\mathcal{F}^{\mathcal{Q}}(K,C,B)$ is a complete minor class for any sets B and C.

SECTION 12: THE MINOR CLASS OF MATROIDS AND RELATED MINOR CLASSES

There are many ways of describing a matroid [18], for example by independent sets, bases, circuits, rank function, closure operator, flats or hyperplanes, any one way uniquely determining the others. It is most convenient here to describe a matroid in terms of its closure operator. Let Q be a finite set. A *closure operator*, on ground set Q, is a function $\sigma:2^Q\to2^Q$ satisfying the following three conditions [18].

(C1) $P\subseteq\sigma(P)$ for all $P\subseteq Q$.

(C2) $P\subseteq R$ implies $\sigma(P)\subseteq\sigma(R)$ for all $P,R\subseteq Q$.

(C3) $\sigma(P)=\sigma(\sigma(P))$ for all $P\subseteq Q$.

Now σ is the closure operator of a matroid if the following condition also holds.

(C4) If $q\notin\sigma(P)$ and $q\in\sigma(P\cup\{r\})$ then $r\in\sigma(P\cup\{q\})$ for all $P\subseteq Q$ and all distinct $q,r\in Q-P$.

In this case it is convenient to say that σ <u>is</u> a *matroid*, on ground set Q.

The minor class $\mathscr{M}$ has functions $\sigma:2^Q\to2^Q$, satisfying (C1), as its structures on ground set Q, finite sets as ground sets, and two manners of point removal called *deletion* and *contraction*. Let $\sigma:2^Q\to2^Q$, with ground set Q, be a structure in $\mathscr{M}$. For $q\in Q$, let $\sigma\backslash q$ denote the *deletion* of q from *s* and let σ/q denote the *contraction* of q from σ. Define $(\sigma\backslash q):2^{Q-\{q\}}\to2^{Q-\{q\}}$ by setting $(\sigma\backslash q)(P)=\sigma(P)-\{q\}$ for all $P\subseteq Q-\{q\}$. Define $(\sigma/q):2^{Q-\{q\}}\to2^{Q-\{q\}}$ by setting $(\sigma/q)(P)=\sigma(P\cup\{q\})-\{q\}$ for all $P\subseteq Q-\{q\}$. Structure isomorphism is defined naturally. It is routine to verify that $\mathscr{M}$ is indeed a minor class. (The proof uses the same reasoning as that for the well known result that matroids form a minor class.)

For any $R\subseteq Q$, deleting (or contracting) R from σ is the same as deleting (or contracting) all the elements of R from σ (in any order). Thus $(\sigma\backslash R):2^{Q-R}\to2^{Q-R}$ is defined by $(\sigma\backslash R)(P)=\sigma(P)-R$ and $(\sigma/R):2^{Q-R}\to2^{Q-R}$ is defined by $(\sigma/R)(P)=\sigma(P\cup R)-R$, for all $P\subseteq Q-R$. Let $\sigma\cdot R$ denote $\sigma\backslash(Q-R)$.

Now $\mathcal{M}^{-}$ has a unique 0-point structure since there is only one function from 2^Q to 2^Q (and this satisfies (C1)). Let $\sigma:2^{\{q\}}\rightarrow 2^{\{q\}}$, on 1-element ground set $\{q\}$, be a structure in $\mathcal{M}^{-}$. Now $\sigma(\{q\})$ must be $\{q\}$, by (C1), but $\sigma(\emptyset)$ could be either $\emptyset$ or $\{q\}$. If $\sigma(\emptyset)=\emptyset$ then σ is called coloop(q), and if $\sigma(\emptyset)=\{q\}$ then σ is called loop(q). Clearly loop(q) and coloop(q) are ψ-equivalent, and hence ψ-structures of $\mathcal{M}^{-}$. (Note also that they both satisfy conditions (C1)-(C4).)

Consider a structure $\sigma:2^Q\rightarrow 2^Q$, on ground set Q, in $\mathcal{M}^{-}$. Every 1-point minor of σ is of the form $(\sigma/P)\cdot q$ (that is, all points in P are contracted, and every other point except q is deleted) for some $P\subseteq Q$ and $q\in Q-P$. By the above definitions it follows that, for all $P\subseteq Q$ and every $q\in Q-P$,

$$\begin{cases} q\in\sigma(P) \text{ exactly when } (\sigma/P)\cdot q=\mathrm{loop}(q) \\ q\notin\sigma(P) \text{ exactly when } (\sigma/P)\cdot q=\mathrm{coloop}(q). \end{cases}$$

Clearly σ is uniquely determined by specifying whether or not $q\in\sigma(P)$ for all $P\subseteq Q$ and $q\in Q-P$ (since by (C1), $\sigma(P)=P\cup\{q\,|\,q\in Q-P \text{ and } q\in\sigma(P)\}$). Therefore σ is uniquely determined by specifying $(\sigma/P)\cdot q$ for all $P\subseteq Q$ and every $q\in Q-P$, that is, σ is uniquely determined by its 1-point minors. Let $\mathcal{M}^{\psi}$ be the sub minor class of $\mathcal{M}^{-}$ consisting of the 0-point and 1-point structures of $\mathcal{M}^{-}$. Then $\mathcal{M}^{\psi}$ is a big subalgebra of $\mathcal{M}^{-}$, and since the 1-point structures are ψ-structures, it is the core of $\mathcal{M}^{-}$ (and there are no other ψ-structures). Also the choice as to whether $q\in\sigma(P)$ or $q\notin\sigma(P)$, can be made independently for all $P\subseteq Q$ and $q\in Q-P$ (since $\sigma:2^Q\rightarrow 2^Q$, where for all $P\subseteq Q$ it holds that $\sigma(P)=P\cup\{q\,|\,q\in Q-P \text{ and } q\in\sigma(P)\}$, satisfies all these statements as well as (C1)). Equivalently the structures in $\mathcal{M}^{-}$ can have their 1-point minors specified arbitrarily so that $\mathcal{M}^{-}$ is a complete minor class, namely the completion of $\mathcal{M}^{\psi}$.

The *minor class of closure operators* $\mathcal{M}^{C}$ is a sub minor class of $\mathcal{M}^{-}$, consisting of exactly those structures satisfying (C1)–(C3), namely closure operators. The *minor class of matroids* $\mathcal{M}$ is a sub minor class of $\mathcal{M}^{-}$ and $\mathcal{M}^{C}$, consisting of exactly those structures satisfying (C1)–(C4), namely matroids. It is verified below that $\mathcal{M}^{C}$ and $\mathcal{M}$ are indeed minor classes, with $\mathcal{M}$ a sub minor class of $\mathcal{M}^{C}$, and $\mathcal{M}^{C}$ a sub minor class of $\mathcal{M}^{-}$. Also $\mathcal{M}^{\psi}$ is a sub minor class of $\mathcal{M}$ and $\mathcal{M}^{C}$ so that $\mathcal{M}^{\psi}$ is the core of $\mathcal{M}$ and $\mathcal{M}^{C}$ and $\mathcal{M}^{-}$ is the completion of $\mathcal{M}$ and $\mathcal{M}^{C}$. It remains to find the natural excluded minors of $\mathcal{M}$ and $\mathcal{M}^{C}$, namely their excluded minors in $\mathcal{M}^{-}$.

As discussed in section 11, the structures in $\mathcal{M}^{-}$ on ground set Q, can be depicted as Q–hypercubes with a "pattern" or "label" on the 1–faces (1–dimensional "edges") indicating whether the corresponding minor is a loop or a coloop. (The rest of the hypercube remains blank.) It is instructive to look at the 2–point structures in $\mathcal{M}^{-}$, say those on the 2–element ground set $\{q,r\}$. These are depicted by squares in $\mathbb{R}^{\{q,r\}}$ with their four 1–dimentional edges (corresponding to the 1–point minors obtainable by removing one of two points in one of two manners) labelled by either L for loop or C for coloop. The labelling scheme is given for a structure σ on ground set $\{q,r\}$ in figure 1. Note that the only possible automorphisms of σ, and the corresponding depiction, are that which fixes q and r (and fixes the corresponding square) and that which swaps q and r (and reflects the square along the dotted line, as in figure 1). No other symmetries of the square correspond to structure automorphisms.

There are of course, 2^4=16 structures on ground set $\{q,r\}$ in $\mathcal{M}^{-}$, but only 10 up to isomorphism. These 10 have been named 2a,2b,$\cdots$,2j as in figure 1. (Observe that 2a,2e,2h and 2j have the bijection which swaps q and r, as an automorphism.) By observation, only 2a,2e,2f and 2j are matroids, so that 2b,2c,2d,2g,2h and 2i are natural excluded minors of $\mathcal{M}$. Also, only 2a,2e,2f,2g and 2j are closure operators, so that 2b,2c,2d,2h and 2i are natural

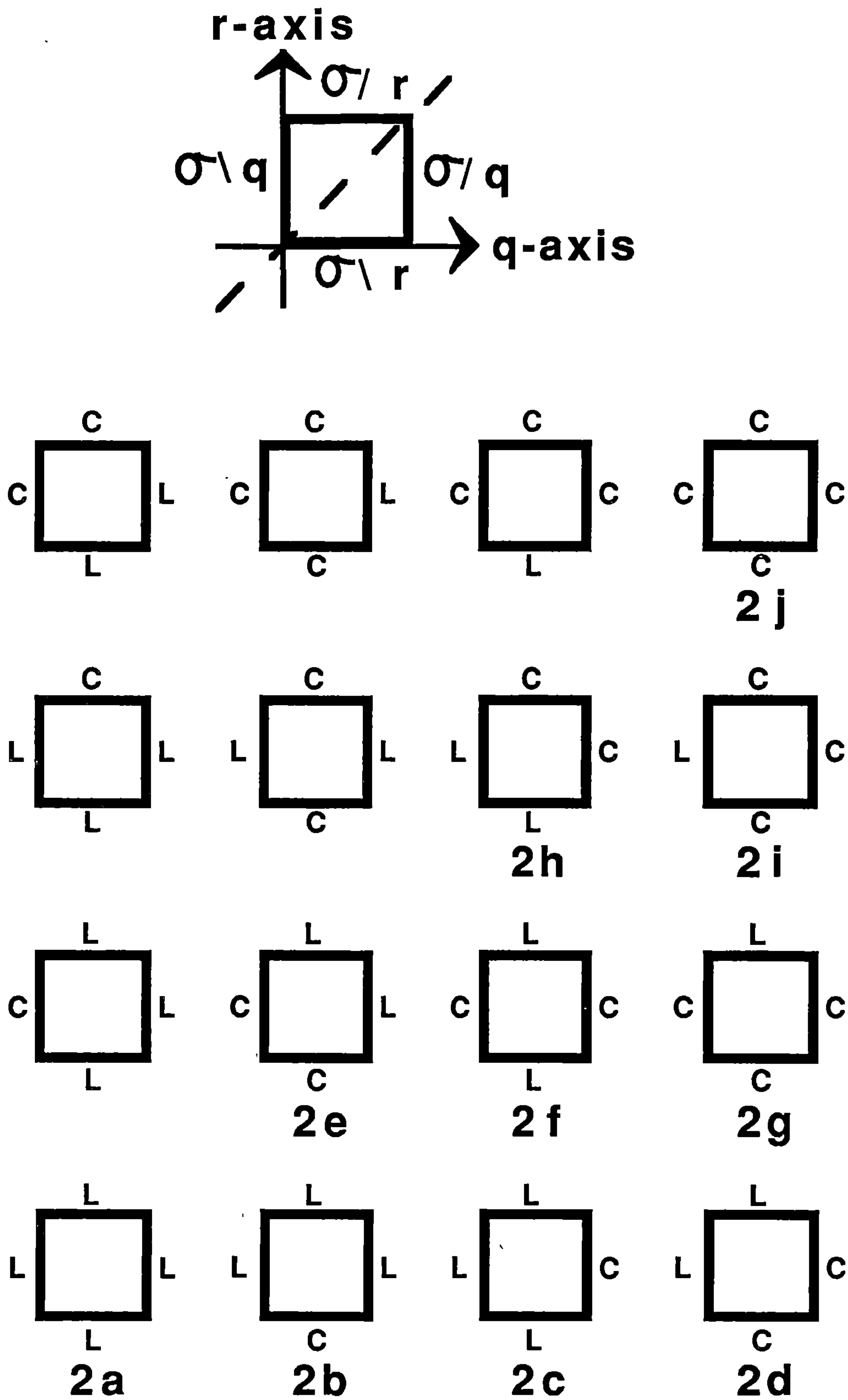

Figure 1: The 2–point structures of $\mathscr{M}$.

excluded minors of $\mathscr{M}^C$. It turns out that these are the only ones, as is now shown.

First suppose that $\sigma:2^Q\to 2^Q$, with ground set Q, satisfies (C1), but not (C2). Then there exist P,R$\subseteq$Q with P$\subseteq$R such that $\sigma(P)\nsubseteq\sigma(R)$. This means that there exists r$\in$Q such that r$\in\sigma(P)$, whereas r$\notin\sigma(R)$. Without loss of generality, we can assume that P is a maximal subset of R with this property. Clearly P$\neq$R, so that there exists q$\in$R–P, and hence $\sigma(P\cup\{q\})\subseteq\sigma(R)$ so that r$\notin\sigma(P\cup\{q\})$. (This shows that q$\neq$r.) Let structure $\rho=(\sigma/P)\cdot\{q,r\}$. Therefore $\rho\backslash q$=loop(r), since r$\in\sigma(P)$, and ρ/q=coloop(r), since r$\notin\sigma(P\cup\{q\})$. It follows that ρ is one of 2c,2d,2h or 2i (see figure 1) and any structure in $\mathscr{M}^-$ which does not satisfy (C2) has one of these as an isominor. Conversely, by reversing the above arguments, it follows that any structure in $\mathscr{M}^-$, having one of 2c,2d,2h or 2i as a minor, fails to satisfy (C2). So the (unnamed) minor class of structures in $\mathscr{M}^-$ which satisfy (C2) has ψ-structures, loop and coloop, and natural excluded minors 2c,2d,2h and 2i.

Now suppose that $\sigma:2^Q\to 2^Q$, with ground set Q, satisfies (C1) and (C2), but not (C3). Then there exists P$\subseteq$Q such that $\sigma(P)\neq\sigma(\sigma(P))$. By (C1) and (C2) it follows that $\sigma(P)\subseteq\sigma(\sigma(P))$ and hence $\sigma(P)$ is a proper subset of $\sigma(\sigma(P))$. Put R=$\sigma(P)$ and, without loss of generality, assume that P is a maximal subset of R, having the property that R=$\sigma(P)$. Clearly P$\neq$R , so that there exists r$\in$R–P, and in particular r$\in$R=$\sigma(P)$. Now $\sigma(P)\subseteq\sigma(P\cup\{r\})\subseteq\sigma(R)$ by (C2) and $\sigma(P)\neq\sigma(P\cup\{r\})$ by the maximality of P. Hence there exists q$\in\sigma(R)$–R such that q$\in\sigma(P\cup\{r\})$. But q$\notin$R=$\sigma(P)$. Let $\rho=(\sigma/P)\cdot\{q,r\}$. (Clearly q$\neq$r, since r$\in$R and q$\notin$R.) Now $\rho\backslash q$=loop(r), since r$\in\sigma(P)$, ρ/r=loop(q), since q$\in\sigma(P\cup\{r\})$, and $\rho\backslash r$=coloop(q), since q$\notin\sigma(P)$. Of the two possibilities for ρ, namely 2b and 2d, the latter does not satisfy (C2), so that ρ must be 2b, and any structure in $\mathscr{M}^-$ satisfying (C1) and (C2), but not (C3), has 2b as a isominor. Reversing the above arguments yields the converse, namely that any structure in $\mathcal{M}^{-}$ satisfying (C1) and (C2) and having 2b as a minor, fails to satisfy (C3). It follows that the ψ–structures of $\mathcal{M}^{C}$ are loop and coloop, and the natural excluded minors are 2b,2c,2d,2h and 2i.

Finally, suppose $\sigma:2^{Q}\to 2^{Q}$, with ground set Q, satisfies (C1),(C2) and (C3) but not (C4), that is, σ is a closure operator but not a matroid. Then there exists $P\subseteq Q$ and distinct $p,q\in Q-P$ such that $q\notin\sigma(P)$ and $q\in\sigma(P\cup\{r\})$ and $r\notin\sigma(P\cup\{q\})$. Let $\rho=(\rho/P)\cdot\{q,r\}$. Then $\rho\backslash r=\text{coloop}(q)$, $\rho/r=\text{loop}(q)$ and $\rho/q=\text{coloop}(r)$. Of the two possibilities for ρ, namely 2g and 2d, the latter does not satisfy (C2) (nor (C3)) so that ρ must be 2g, and any structure in $\mathcal{M}^{C}-\mathcal{M}$ has 2g as a minor. Reversing the above arguments yields the converse, namely that any structure in $\mathcal{M}^{C}$, which has 2g as a minor, is not in $\mathcal{M}$. It follows that 2g is the only excluded minor of $\mathcal{M}$ in $\mathcal{M}^{C}$, and the ψ–structures of $\mathcal{M}$ are loop and coloop, and the natural excluded minors are 2b,2c,2d,2g,2h and 2i. Therefore a matroid on ground set Q, can be depicted as a patterned Q–hypercube with 1–faces (edges) given one of two labels corresponding to loop and coloop, such that none of the square faces looks like 2b,2c,2d,2g,2h or 2i, as in figure 1.

The natural excluded minors of any sub minor class of $\mathcal{M}$ can be found from its excluded minors in $\mathcal{M}$, and visa versa, as shown in section 11. If the sub minor class contains loop and coloop (making it non–trivial) then this involves merely adding to its excluded minors in $\mathcal{M}$, the six natural excluded minors of $\mathcal{M}$. In particular, a sub minor class of $\mathcal{M}$ has finitely many excluded minors in $\mathcal{M}$, if and only if it has finitely many natural excluded minors.

There are two minor class automorphisms of $\mathcal{M}^{\psi}$ (and hence $\mathcal{M}^{-}$) since loop and coloop can either be fixed or swapped. These have four minor class mixed automorphisms, since deletion and contraction can either be fixed or swapped. By considering their natural excluded minors, $\mathcal{M}$ and $\mathcal{M}^{C}$ have only the identity as an automorphism, but $\mathscr{M}$ also has the mixed automorphism which swaps loop and coloop, and swaps deletion and contraction. This sends each matroid σ, to its *dual*, denoted σ^*. For any matroid $\sigma \in \mathscr{M}$ and any point $q \in G(\sigma)$, it follows that $(\sigma^*)^* = \sigma$, $(\sigma \backslash q)^* = (\sigma^*)/q$ and $(\sigma/q)^* = (\sigma^*)\backslash q$. This mixed automorphism permutes the natural excluded minors of $\mathscr{M}$, since it swaps 2b with 2g, 2c with 2i, and fixes 2d and 2h. The patterned hypercube depicting the dual σ^*, is obtained from that depicting the matroid σ, by sending each point $(x_q \mid q \in Q)$ of the Q–hypercube, to $(1-x_q \mid q \in Q)$ (since deletion and contraction are swapped) and changing each 1–face labelled loop to coloop, and visa versa.

Let $\mathscr{S}$ be a minor class with the same ground sets and manners of point removal as $\mathscr{M}$ (so that $\mathscr{S}$ is in the same variety as $\mathscr{M}$). Any homomorphism $\alpha: \mathscr{S} \to \overline{\mathscr{M}}$ (which may send all of $\mathscr{S}$ into $\mathscr{M}$ or $\mathscr{M}^C$) is uniquely determined by specifying $\alpha|^{\mathscr{M}^{\psi}}$ (see section 10 or 11) or equivalently, by specifying which structures on a 1–element ground set $\{q\}$, say, are sent to loop (q), and which are sent to coloop (q). Conversely, from sections 10 and 11, this latter specification can be made arbitrarily, and there always exists, uniquely of course, such a homomorphism $\alpha: \mathscr{S} \to \overline{\mathscr{M}}$. While this only guarantees that α sends $\mathscr{S}$ into $\overline{\mathscr{M}}$, it can be checked that α sends $\mathscr{S}$ into $\mathscr{M}$ (or $\mathscr{M}^C$) by checking that no structure of $\mathscr{S}$ is sent by α, to a natural excluded minor of $\mathscr{M}$ (or $\mathscr{M}^C$). This is easy in practice, since these only have order 2.

Recall the minor class $\mathscr{W}$ defined in section 3 and $\mathscr{W}/q$ (a particular homomorphic image of $\mathscr{W}$) defined in section 5. From the fact that $\mathscr{W}$ is isomorphic to $\mathscr{T}^{\mathscr{Q}}(K,\emptyset,B)$ for $|B|=2$ and the appropriate $\mathscr{Q}$ and K, and the discussion about $\mathscr{T}^{\mathscr{Q}}(K,\emptyset,B)$ in the previous section, it follows that $\mathscr{W}$ has two ψ–equivalent 0–point structures and no natural excluded minors. It turns out that $\mathscr{W}/q$ has two equivalent 1–point structures, which can be called loop and coloop, and four natural excluded minors, namely 2b,2c,2g and 2i. (The proof is left to the reader.) It follows that $\mathcal{M}$ can be embedded in $\mathcal{W}/q$ with the exclusion of two further minors, namely 2d and 2h. Observe that this is an example of the embedding theorem 8.4.

SECTION 13: MINOR CLASSES OF LINEAR DEPENDENCIES

Coordinatisability over a ring

If A is an algebra and $B \subseteq A$, then it is well known that the function from 2^B to 2^B, which sends each $C \subseteq B$ to $B \cap (\mathrm{Ainc}C)$, is a closure operator. For various kinds of algebras, there is interest in what such closure operators look like. In particular, matroid theorists are interested in this problem for vector spaces over a fixed field. We specialise to this case after starting with the case of unital modules over a fixed ring with (multiplicative) identity.
(A straightforward agrument shows that this is no less general than modules over rings, in the context of this discussion.)

Let R be a ring with multiplicative identity $1 \in R$. A (*left*) *unital module* M, *over* R, consists of an abelian group, also called M for convenience, together with *scalar multiplication*, multiplying elements of M by "scalars" from R (on the left), satisfying the following rule. If $a,b \in R$ and $m,n \in M$, then $(a+b)m = am+bm$, $a(m+n) = am+an$, $(ab)m = a(bm)$, and $1m = m$. If R is a field then, as is well known, M is a vector space over R. From now on, a left unital module will be simply called a module.

Consider a finite set Q and a function $f:Q \to M$. If $P \subseteq Q$ and $q \in Q$, then q is *dependent* (in f) on P exactly when f(q) can be expressed in terms of all $f(p) \in M$ where $p \in P$, using the operations of the module. Equivalently, q is dependent on P, exactly when f(q) is equal to a linear combination $\sum_{p \in P} a_p f(p)$ for some family $(a_p \mid p \in P)$ of elements of R. Define a function $\sigma_f : 2^Q \to 2^Q$ such that $q \in \sigma(P)$ exactly when q is dependent (in f) on P, for every $P \subseteq Q$ and every $q \in Q$. Trivially, if $q \in P$, then q is dependent on P, so that σ_f satisfies (C1) and $\sigma_f \in \mathscr{M}$. Observe that σ_f could be equivalently defined by stating that σ_f satisfies (C1) and considering only $q \in Q-P$.

A closure operator of the form σ_f, where $f:Q \to M$ is a function, Q is a ground set, and M is a module over R, is *coordinatisable over* R. (This coincides with the usual matroid definition when R is a field or division ring.)

Consider R^Q, the set of all "vectors" $(a_q|q\in Q)$ where $a_q\in R$ for every $q\in Q$. Then R^Q is a module over R if addition and scalar multiplication are defined component wise. That is, for all "vectors" $a=(a_q|q\in Q)$ and $b=(b_q|q J Q)$ in R^Q and all "scalars" $c\in R$, let $a+b=(a_q+b_q|q\in Q)$ and $ca=(ca_q|q\in Q)$. Any subset of R^Q is a submodule of R^Q, if it is closed under addition and scalar multiplication. For the above function $f:Q\to M$, let $S_f\subseteq R^Q$ be defined by $(a_q|q\in Q)\in S$ if and only if $\sum_{q\in Q} a_q f(q)=0$. Thus S_f expresses which linear combinations of the f(q), over $q\in Q$, give zero. Clearly S_f is a submodule of R^Q. Conversely, for any submodule S of R^Q, there is a module M, over R, and a function $f:Q\to M$, such that $S=S_f$, as the following construction shows. For each $a\in R^Q$, let $a+S=\{a+s|s\in S\}$. The factor module R^Q/S has elements a+S for all $a\in R^Q$. For all $a,b\in R^Q$ and $c\in R$ it holds that $(a+S)+(b+S)=(a+b)+S$ and $c(a+S)=(ca)+S$. It is well known that this is well defined exactly when S is a submodule of R^Q. For each $t\in Q$, let $1^t=(1^t_q|q\in Q)$ where $1^t_q=1$ if t=q and 0 otherwise. Define $f:Q\to R^Q/S$ by setting $f(t)=1^t+S$ for each $t\in Q$. Then $(a_t|t\in Q)\in S_f$ means that $\sum_{t\in Q} a_t f(t)=0$. That is $(a_t|t\in Q)\in S_f$ exactly when $0+S=\sum_{t\in Q} a_t((1^t_q|q\in Q)+S)=(a_q|q\in Q)+S$, or equivalently $(a_t|t\in Q)\in S$, as required.

Let $a=(a_q|q\in Q)$ be an element of R^Q. Then the *support* of a, denoted supp(a), is equal to $\{q|q\in Q \text{ and } a_q\neq 0\}$. For any submodule S of R^Q define a function $\sigma_S:2^Q\to 2^Q$, satisfying (C1), as follows. If $P\subseteq Q$ and $q\in Q-P$, then $q\in\sigma_S(P)$ exactly when there exists $a=(a_t|t\in Q)\in S$ with $\text{supp}(a)\subseteq P\cup\{q\}$ and $a_q=1$. If routine to check, for any function $f:Q\to M$ where M is a module over R, that $\sigma_{S_f}=\sigma_f$. Since every submodule S, of R^Q, is of the form S_f, it follows that a closure operator, on ground set Q, is coordinatisable over R, if and only if it is of the form σ_S for some submodule S of R^Q. So in studying such closure operators, we need not consider functions of the form $f:Q \to M$ at all. This is essentially the approach of Tutte [15] when R is a finite field of the integers, and all the submodules of R^Q for each ground set Q, are called chain groups. (They are also called linear codes by coding theorists.)

The Minor Class $\mathscr{D}(R)$

The minor class $\mathscr{D}(R)$ of (*linear*) *dependencies over* R, has finite ground sets, the submodules of R^Q as its structures on ground set Q, and two manners of point removal called *deletion* and *contraction*. Let S, on ground set Q, be a submodule of R^Q. As usual, denote the deletion and contraction of each $q \in Q$ from S by $S \backslash q$ and S/q respectively. Now $(S \backslash q) \subseteq R^{Q-\{q\}}$ is defined by "intersecting" S with $R^{Q-\{q\}}$, that is, the elements of $S \backslash q$ are $(a_t \mid t \in Q-\{q\}) \in R^{Q-\{q\}}$ for which there exists $(a_t \mid t \in Q) \in S$ with $a_q=0$. And $(S/q) \subseteq R^{Q-\{q\}}$ is defined by "projecting" S onto $R^{Q-\{q\}}$, that is, the elements of S/q are $(a_t \mid t \in Q-\{q\}) \in R^{Q-\{q\}}$ for which there exists $(a_t \mid t \in Q) \in S$, (regardless of the value of a_q). Define structure isomorphism naturally. It is routine to show that $\mathscr{D}(R)$ is indeed a minor class. Also, with S as above, and $P,N \subseteq Q$ where $P \cap N=\emptyset$, the submodule $S/P \backslash N$ of $R^{Q-(P \cup N)}$ (on ground set $Q-(P \cup N)$) is obtained from S by contracting all points in P and deleting all points in N. The elements of $S/P \backslash N$ are $(a_q \mid q \in Q-(P \cup N)) \in R^{Q-(P \cup N)}$ for which there exists $(a_q \mid q \in Q) \in S$ with $\mathrm{supp}((a_q \mid q \in Q)) \subseteq Q-N$ (that is, $a_q=0$ for all $q \in N$). As usual let $S \cdot N$ denote $S \backslash (Q-N)$.

The Homomorphism $\alpha: \mathscr{D}(R) \to \mathscr{M}^C$

Let $\alpha: \mathscr{D}(R) \to \mathscr{M}^{-}$ be the sort (ground set) respecting function which sends each $S \in \mathscr{D}(R)$, on ground set Q, to $\sigma_S \in \mathscr{M}^{-}$. We shall show that α is a homomorphism. Certainly $\alpha|^{\mathscr{M}^{\psi}}$ is a homomorphism, which extends uniquely to a homomorphism from $\mathscr{D}(R)$ to $\mathscr{M}$, which would have to be α, if α is to be a homomorphism. (See section 12.) Consider a ground set Q, a submodule S of R^Q, a subset $P \subseteq Q$, and a point $q \in Q-P$. Now $((\alpha(S))/P)\cdot q = \text{loop}(q)$ if and only if $q \in (\alpha(S))(P)$. This means that there exists $a=(a_t | t \in Q) \in S$ with $\text{supp}(a) \subseteq P \cup \{q\}$ and $a_q=1$, or equivalently there exists $(a_t | t \in \{q\}) \in (S/P)\cdot q$ with $a_q=1$, which means that α sends $(S/P)\cdot q$ to loop(q). Therefore $((\alpha(S))/P)\cdot q = \alpha((S/P)\cdot q)$ (since there are only two possible values) and by the arbitrariness of S, P and q, it follows that α is the unique homomorphism from $\mathscr{D}(R)$ to $\mathscr{M}$ such that $\alpha|^{\mathscr{M}^{\psi}} \leq \alpha$. In particular α is a homomorphism.

It is instructive to examine the 1–point structures in $\mathscr{D}(R)$, and the effect of α on these, since this uniquely determines α (see section 12). If $\{q\}$ is a 1–element ground set then the structures in $\mathscr{D}(R)$ with ground set $\{q\}$ are the submodules of $R^{\{q\}}$. There is an obvious correspondence between $R^{\{q\}}$ and R, and the submodules of $R^{\{q\}}$ correspond with the left ideals of R (or ideals of R, if R is commutative), and for convenience we treat the 1–point structures as left ideals of R. For any left ideal S of R, it holds that $\alpha(S)=\text{loop}$ if and only if $1 \in S$, or equivalently $S=R$. So the only submodule (or left ideal) of $R^{\{q\}}$ (or R), which α sends to loop(q), is $R^{\{q\}}$ (or R) and the rest are sent to coloop(q). When R is a field, the only ideals of R are $\{0\}$ and R, so that the only structures in $\mathscr{D}(R)$, on ground set $\{q\}$, are $\{0\}^{\{q\}}$ and $R^{\{q\}}$, and these shall be called coloop(q) and loop(q) respectively (so that α sends loop to loop and coloop to coloop).

As stated earlier, but not yet proved here, $\alpha(S)=\sigma_S$ is a closure operator for every $S \in \mathscr{D}(R)$. To check this, it is sufficient to show that α sends no element of $\mathscr{D}(R)$ to a natural excluded minor of $\mathscr{M}^C$. Suppose $S \in \mathscr{D}(R)$ is such a structure. Since all the natural excluded minors of $\mathscr{M}^C$ have order 2, so must S. Hence, without loss of generality, S is a submodule of $R^{\{q,r\}}$ for some 2-element ground set {q,r}. Write the elements of $R^{\{q,r\}}$ as ordered pairs, with the first component labelled by q, and the second by r. Suppose $\alpha(S)\backslash r=\mathrm{loop}(q)$. Then $\alpha(S\backslash r)=\mathrm{loop}(q)$ so that $S\backslash r$ is all of $R^{\{q\}}$. By the definition of deletion and contraction in $\mathscr{D}(R)$, it follows that $(a,0)\in S$ for every $a\in R$ and hence S/r is all of $R^{\{q\}}$, so that $\alpha(S)/r=\alpha(S/r)=\mathrm{loop}(q)$. Therefore $\alpha(S)$ cannot be 2c,2d,2h or 2i (see figure 1). Instead now, suppose that $\alpha(S)\backslash r=\mathrm{loop}(q)$ and $\alpha(S)/q=\mathrm{loop}(r)$. Thus $\alpha(S\backslash r)=\mathrm{loop}(q)$ so that $S\backslash r$ is all of $R^{\{q\}}$, and $\alpha(S/q)=\mathrm{loop}(q)$ so that S/q is all of $R^{\{r\}}$. Therefore, for any $b\in R$ there exists $a\in R$ such that $(a,b)\in S$, but also $(a,0)\in S$ so that $(a,b)-(a,0)=(0,b)\in S$ and hence $S\backslash q$ is all of $R^{\{r\}}$, so that $\alpha(S)\backslash q=\alpha(S\backslash q)=\mathrm{loop}(q)$. This shows that $\alpha(S)$ cannot be 2b (see figure 1). Therefore α does indeed send all of $\mathscr{D}(R)$ into $\mathscr{M}^C$, so that $\alpha:\mathscr{D}(R)\to\mathscr{M}^C$ is well defined. Although this is well known, the method of proof is different.

To see whether or not α sends $\mathscr{D}(R)$ into $\mathscr{M}$, it suffices to check whether or not α sends any $S\in\mathscr{D}(R)$ on ground set {q,r}, to 2g (see figure 1). If R is not a division ring, then there exists $a\in R$ such that $ca\neq 1$ for every $c\in R$. Let S be the submodule generated by (1,a), that is, $S=\{(c,ca)\,|\,c\in R\}$. Thus S/r is all of $R^{\{q\}}$ and $\alpha(S)/r=\alpha(S/r)=\mathrm{loop}(q)$. Now $(c,1)\notin S$ for any $c\in R$, in particular $(0,1)\notin S$, so that neither $S\backslash q$ nor S/q are all of $R^{\{r\}}$. Hence $\alpha(S)\backslash q=\alpha(S\backslash q)=\mathrm{coloop}(r)$ and $\alpha(S)/q=\alpha(S/q)=\mathrm{coloop}(r)$. Also $(1,0)\notin S$ so that $S\backslash r$ is not all of $R^{\{q\}}$ and $\alpha(S)\backslash r=\alpha(S\backslash r)=\mathrm{coloop}(q)$. This shows that $\alpha(S)$ is 2g (see figure 1), a natural excluded minor of $\mathscr{M}$, and therefore α does not send all of $\mathscr{D}(R)$ into $\mathscr{M}$. Conversely, suppose that R is a division ring. Suppose that $S\in\mathscr{D}(R)$ on ground set {q,r} is such that $\alpha(S)/r=\mathrm{loop}(q)$ and $\alpha(S)\backslash r=\mathrm{coloop}(q)$. Then there exists $a\in R$ such that $(a,0)\notin S$ and so there exists $b\in R$ with $b\neq 0$ such that $(a,b)\in S$. But then for any $c\in R$ it follows that $cb^{-1}(a,b)=(cb^{-1}a,c)\in S$. So S/q is all of $R^{\{r\}}$ and $\alpha(S)/q=\alpha(S/q)=\mathrm{loop}(q)$.

Therefore α(S) cannot be 2g so that α does send $\mathscr{D}$(R) into $\mathscr{M}$ in this case. Thus α: $\mathscr{D}$(R)$\to\mathscr{M}$ is well defined when R is a division ring.

Some Matroid Theory

The ψ-description of $\mathscr{D}$(R), when R is a field, is given later, but first some preliminary matroid theory is needed. While the definitions and statements below are standard matroid theory [18], they are given in a minor class theoretic fashion. Matroids are taken to be the elements of the minor class $\mathscr{M}$ with ψ-structures loop and coloop and natural excluded minors 2b,2c,2d,2g,2h and 2i (see figure 1), and the only facts used about matroids are that they are uniquely determined by the 1-point minors, and that they do not have as a minor, any of the six natural excluded minors. (Any isomorphic copy of $\mathscr{M}$ could be used, since it is the algebraic structure of $\mathscr{M}$ which is important, not its elements.)

For any ground set Q, if m and n are integers with $0\leq m\leq n=|Q|$, then let $U^n_m(Q)$ be the matroid on ground set Q such that for every $P\subseteq Q$ and every $q\in Q-P$ it holds that $(U^n_m(Q)/P)\cdot q$=coloop(q) exactly when $|P|<m$ and loop(q) exactly when $|P|\geq m$. The matroid $U^n_m(Q)$ is called the *uniform matroid of rank* m *and order* n (*on ground set* Q). (Observe that the dual of $U^n_m(Q)$ is $U^n_{n-m}(Q)$.) In particular coloop(q)=$U^1_1(q)$ and loop(q)=$U^1_0(q)$. (Brackets are omitted when this is unambiguous. For example $U^3_2(\{p,q,r\})$ is denoted $U^3_2(p,q,r)$.)

The direct sum $\rho+\sigma$ of two matroids ρ and σ with disjoint ground sets (that is $G(\rho)\cap G(\sigma)=\emptyset$) is a matroid with ground set $G(\rho)\cup G(\sigma)$, with 1-point minors (uniquely determining $\rho+\sigma$) as follows. If $P\subseteq Q$ and $q\in Q-P$, let $((\rho+\sigma)/P)\cdot q$ be $(\rho/(P\cap G(\rho)))\cdot q$ whenever $q\in G(\rho)$ and $(\sigma/(P\cap G(\sigma)))\cdot q$ whenever $q\in G(\sigma)$. Observe that $(\rho+\sigma)^*=\rho^*+\sigma^*$. For example, the matroid 2f (see figure 1) is $U^1_0(q)+U^1_1(r)$. The direct sum of two matroids is a matroid, since none of the six natural excluded is the direct sum of two matroids. The three matroids 2a,2e and 2j (see figure 1) are $U_0^2(q,r)$, $U_1^2(q,r)$ and $U_2^2(q,r)$ respectively.

Consider the minor class $(\mathcal{M}\times\mathcal{M})\mathrm{exc}((U_1^1(q),U_0^1(q)))$ for some 1–element ground set $\{q\}$. Its structures are pairs (ρ,σ) of matroids ρ and σ (on the same ground set Q, say) such that $(U_1^1(q),U_0^1(q))$ is not an isominor of (ρ,σ). Equivalently, for any $P\subseteq Q$ and any $q\in Q-P$, it is not the case that $(\rho/P)\cdot q=U_1^1(q)$ and $(\sigma/P)\cdot q=U_0^1(q)$. Define a relation $\prec$ on matroids where $\rho\prec\sigma$ if and only if $(\rho,\sigma)\in(\mathcal{M}\times\mathcal{M})\mathrm{exc}((U_1^1(q),U_0^1(q)))$. (In particular $G(\rho)=G(\sigma)$.) The fact that $\prec$ is a partial order on matroids follows from the fact that it is a partial order on 1–point matroids. (Matroid theorists may recognise this as the strong map partial order [9].) If σ is a matroid on ground set Q and $q\in Q$, then the exclusion of 2c,2d,2h and 2i (see figure 1) implies that $(\sigma/q)\prec(\sigma\backslash q)$. If $M\subseteq N\subseteq Q$ and $P\subseteq Q-N$, it follows by induction that $((\sigma/N)\cdot P)\prec((\sigma/M)\cdot P)$.

In particular, if $q\in Q$ and $\sigma/(Q-\{q\})=U_1^1(q)=\mathrm{coloop}(q)$, then for any $P\subseteq Q-\{q\}$ it follows that $(\sigma/P)\cdot q=\mathrm{coloop}(q)$. In this case q *is a coloop of* σ. By the exclusion of 2g and 2i (see figure 1) it follows that if q is a coloop of σ, then $(\sigma\backslash q)=(\sigma/q)$ and hence $\sigma=(\sigma\backslash q)+\mathrm{coloop}(q)$. If $q\in Q$ is such that $\sigma\backslash(Q-\{q\})=\mathrm{loop}(q)$ then q *is a loop of* σ. (Observe that q is a loop of σ exactly when q is a coloop of σ^*.) In this case $(\sigma/P)\cdot q=\mathrm{loop}(q)$ for every $P\subseteq Q\{q\}$, and by the exclusion of 2b and 2c (see figure 1) it follows that $(\sigma\backslash q)=(\sigma/q)$ and hence $\sigma=(\sigma\backslash q)+\mathrm{loop}(q)$. Conversely if $(\sigma\backslash q)=(\sigma/q)$, then by the exclusion of 2b,2c,2g and 2i (see figure 1) it follows that q is either a loop or coloop of σ.

If q and r are distinct elements of Q, and $\sigma/(Q-\{q,r\})=U_1^2(q,r)$, then q and r are *coparallel* in σ. (Being coparallel in σ is a transitive relation, since if p,q,r are distinct elements of Q with p and q and also q and r being coparallel in σ, then $\sigma/(Q-\{p,q,r\})$ must be $U^3_2(p,q,r)$, by the exclusion of 2h,2i,2g and 2d (see figure 1). Therefore $\sigma/(Q-\{p,r\})=U^3_2(p,q,r)/q=U^2_1(p,r)$ and p and r are coparallel in σ.) If q and r are coparallel in σ, then bijection from Q to Q which fixes $Q-\{q,r\}$ and swaps q and r, is an automorphism of σ, as the following argument shows. If $P\subseteq Q-\{q,r\}$ then $(\sigma/P)\cdot\{q,r\}$ is equal to either $U^2_1(q,r)$ or $U^2_2(q,r)$ (since these are the only matroids that are above $U^2_1(q,r)$ in the partial order $\prec$) and these have the bijection swapping q and r as an automorphism. Also $\sigma/q/r=\sigma/r/q$ and $\sigma\backslash q\backslash r=\sigma\backslash r\backslash q$, as always. Now r is a coloop of σ/q, since $(\sigma/q)/(Q-\{q,r\})=(\sigma/Q-\{q,r\})/q=U^2_1(q,r)/q=\text{coloop}(q)$, and similarly q is a coloop of σ/r. Therefore $\sigma/q\backslash r=\sigma/q/r=\sigma/r/q=\sigma/r\backslash q$. This covers all the 1–point minors of σ, so that σ does indeed have the abovementioned automorphism. If q and r are distinct elements of Q and $\sigma\backslash(Q-\{q,r\})=U^2_1(q,r)$ then q and r are *parallel* in σ. Observe that q and r are parallel in σ exactly when they are coparallel in σ^*. By "dualising" the arguments above, it follows that being parallel in σ, is a transitive relation, and also that if q and r are parallel in σ, then the bijection from Q to Q which fixes $Q-\{q,r\}$ and swaps q and r, is an automorphism of σ.

The *rank* of the matroid σ, denoted $\text{rk}(\sigma)$, is defined inductively as follows. The rank of the 0–point matroid is zero, and for any $q\in Q$, it holds that $\text{rk}(\sigma)$ is either $\text{rk}(\sigma\backslash q)$ in the case that $\sigma/(Q-\{q\})=\text{loop}(q)$ or $\text{rk}(\sigma\backslash q)+1$ in the case that $\sigma/(Q-\{q\})=\text{coloop}\{q\}$. (For example, the rank of $U^n_m(Q)$ is m, and in particular the rank of loop(q) is 0, and the rank of coloop(q) is 1.) The exclusion of 2b,2c,2d,2g and 2i (see figure 1), ensures that this definition is consistent, that is, independent of the choice of q. For any $P\subseteq Q$, the rank of P in σ, denoted $\text{rk}_\sigma(P)$ (or simply rk(P) if σ is known from the context), is defined to be $\text{rk}(\sigma\cdot P)$. If $P\subseteq Q$ and $\text{rk}(P)<\text{rk}(\sigma)$ then by the inductive definition of rank there must exist a 1–point minor of σ/P, say on ground set $\{q\}$, which is coloop(q). But by the properties of the partial order $\prec$, it follows that $(\sigma/P)\cdot q$=coloop(q) and this equals $(\sigma\cdot P\cup\{q\})/P$, so that $rk(P\cup\{q\})=rk(P)+1$. By induction it follows that, for any non-negative integer $m\leq rk(\sigma)$, there exists $P\subseteq Q$ with $rk(P)=m$ and $rk(P)=|P|$.

The *corank*, $rk^*(\sigma)$, of σ, is equal to $rk(\sigma^*)$, the rank of the dual of σ. If $P\subseteq Q$, then the corank of P, denoted $rk^*(P)$, is defined to be $rk(\sigma^*\cdot P)$. By dualising the above argument, it follows that, for any non-negative integer $m\leq rk^*(\sigma)$, there exists $P\subseteq Q$ with $rk^*(P)=m$ and $rk^*(P)=|P|$. An alternative definition of $rk(\sigma)$ and $rk^*(\sigma)$, easily deduced from the inductive definition, is given by considering the Q–hypercube visualisation of σ, although it can be expressed formally. Consider any $|Q|$ edge path from the vertex representing $\sigma\backslash Q$ (namely $(0|q\in Q)$) to the vertex representing σ/Q (namely $(1|q\in Q)$). Then $rk(\sigma)$ (respectively $rk^*(\sigma)$) is the number of edges labelled coloop (respectively loop) in the path. The exclusion of 2b,2c,2d,2g and 2i (see figure 1) ensures that this is independent of the choice of path. It follows that $rk(\sigma)+rk^*(\sigma)=|Q|$, the order of σ.

The ψ–description of $\mathscr{D}(R)$ When R is a Field

From now on we restrict our attention to the case where R is a field. In this case, let us find the ψ–description of $\mathscr{D}(R)$. This task is simplied by the fact that $\mathscr{D}(R)$ has a duality like that of $\mathscr{M}$. Consider a ground set Q and any subspace S of the vector space R^Q. The *dual* of S, denoted S^*, is the submodule of R^Q containing all elements $(b_q|q\in Q)$ such that $\sum_{q\in Q} a_q b_q=0$ for every $(a_q|q\in Q)\in S$. (Also S^* is called the orthogonal subspace of S.) It is routine to check that $\mathscr{D}(R)$ has a mixed automorphism which swaps deletion and contraction, and which swaps each $S\in\mathscr{D}(R)$ with its dual $S^*\in\mathscr{D}(R)$. Also, for any $S\in\mathscr{D}(R)$ and any $q\in G(S)$, it holds that $(S^*)^*=S$ and $(S\backslash q)^*=(S^*)/q$ and $(S/q)^*=(S^*)\backslash q$. The homomorphism $\alpha:\mathscr{D}(R)\to\mathscr{M}$

"respects" duality in the sense that $\alpha(S^*)=(\alpha(S))^*$. The above mixed automorphism sends ψ-structures to ψ-structures, and if it is extended to $\overline{\mathscr{D}(R)}$ (see section 11) then it permutes the natural excluded minors.

Recall the homomorphism $\alpha: \mathscr{D}(R) \to \mathscr{M}$. As discussed in section 12 and before, α extends uniquely to a homomorphism from $\overline{\mathscr{D}(R)}$ to $\mathscr{M}^-$, which can be denoted $\alpha: \overline{\mathscr{D}(R)} \to \mathscr{M}^-$. In particular, α is defined on the natural excluded minors of $\mathscr{D}(R)$. (Note that just because some elements of $\overline{\mathscr{D}(R)}$ are not in $\mathscr{D}(R)$ and some subsets of R^Q are not in $\mathscr{D}(R)$, it does not follow that the structures in $\overline{\mathscr{D}(R)}$ on ground set Q, can be thought of as subsets of R^Q, because they cannot.) The homomorphism α is useful because the 1–point minors of S are the same as those of $\alpha(S)$, so that α gives a convenient means of specifying the 1–point minors of S.

Let $S \in \mathscr{D}(R)$ have ground set Q, so that S is a subspace of R^Q. For any $P \subseteq Q$, identify the points $(a_q \mid q \in Q-P)$ in R^{Q-P}, with the points $(a_q \mid q \in Q)$ in R^Q, such that $a_q=0$ for every $q \in P$. It then follows that $S\backslash q = S \cap R^{Q-\{q\}}$ for any $q \in Q$. Also for any distinct $q,r \in Q$, it follows that

$R^{Q-\{q\}} \cap R^{Q-\{r\}} = R^{Q-\{q,r\}}$ and hence

$(S\backslash q) \cap (S\backslash r) = (S \cap R^{Q-\{q\}}) \cap (S \cap R^{Q-\{r\}}) = S \cap R^{Q-\{q,r\}} = S\backslash q\backslash r$. Let dim(S) denote the *dimension* of S as a vector space. If $\dim(S)=\dim(S\backslash q)$ then $S \subseteq R^{Q-\{q\}}$ so that there is no $a=(a_t \mid t \in Q) \in S$ with $q \in \mathrm{supp}(a)$, and hence $S/Q-\{q\}$ is the 0–dimensional subspace $\{0\}^{\{q\}}$ of $R^{\{q\}}$, namely coloop(q), (a 1–point structure in $\mathscr{D}(R)$). So in this case $\alpha(S)/(Q-\{q\}) = \alpha(S/Q-\{q\}) = \mathrm{coloop}(q)$, (a 1–point structure in $\mathscr{M}$), and in fact q <u>is</u> a coloop of $\alpha(S)$. However if $\dim(S)=\dim(S\backslash q)+1$, then there exists $a=(a_t \mid t \in Q) \in S$ with $q \in \mathrm{supp}(a)$, so that $(a_t \mid t \in \{q\}) \in (S/Q-\{q\})$ with $a_q \neq 0$, and hence $S/Q-\{q\}$ is the 1–dimensional subspace $R^{\{q\}}$ of $R^{\{q\}}$, namely loop(q) (in $\mathscr{D}(R)$). In this case $\alpha(S)/(Q-\{q\}) = \alpha(S/(Q-\{q\}) = \mathrm{loop}(q)$ (in $\mathscr{M}$). Comparing this to the inductive definition of rank, it follows that $\dim(S)=\mathrm{rk}^*(\alpha(S))$, the <u>co</u>rank of $\alpha(S)$.

It is instructive to determine when $S \in \mathscr{D}(R)$, on ground set Q, is uniquely determined by $S\backslash q$ and $S\backslash r$, for some given distinct $q,r \in Q$. That is, if $T \in \mathscr{D}(R)$, on ground set Q, and $S\backslash q = T\backslash q$ and $S\backslash r = T\backslash r$, does it follow that $S=T$? Now $S\backslash q$ (respectively $S\backslash r$) can be treated as subspaces of R^Q with all elements having q^{th} (respectively r^{th}) coordinate being zero. Define the subspace $(S\backslash q)+(S\backslash r)$ (unrelated to direct sums of matroids) to be $\{ca+db \mid c,d \in R$ and $a \in S\backslash q$ and $b \in S\backslash r\}$. One possibility for T is to be $(S\backslash q)+(S\backslash r)$. In any case S (and any T with $S\backslash q = T\backslash q$ and $S\backslash r = T\backslash r$) must have $(S\backslash q)+(S\backslash r)$ as a subspace. The dimension of $(S\backslash q)+(S\backslash r)$ is $\dim(S\backslash q)+\dim(S\backslash r)-\dim(S\backslash q\backslash r)$, since $S\backslash q\backslash r$ is $(S\backslash q)\cap(S\backslash r)$. Now the dimensions of $S\backslash q$ and $S\backslash r$ will each be 0 or 1 more than $\dim(S\backslash q\backslash r)$, and will each be 0 or 1 less than $\dim(S)$. These four differences correspond to the four 1–point minors (0 for coloop and 1 for loop) of the matroid $\alpha(S/(Q-\{q,r\}))=\alpha(S)/(Q-\{q,r\})=\rho$, say, on ground set $\{q,r\}$. There are five such matroids, namely 2a,2e,2f (and the other matroid on $\{q,r\}$, isomorphic to 2f, depicted just above 2e in figure 1) and 2j (or $U_0^2(q,r)$, $U_1^2(q,r)$ $U_0^1(q)+U_1^1(r)$, $U_1^1(q)+U_0^1(r)$ and $U_2^2(q,r)$, respectively). Now S is uniquely determined by $S\backslash q$ and $S\backslash r$ if and only if ρ is uniquely determined by $\rho\backslash q$ and $\rho\backslash r$. (If there is only one possibility for ρ, then there is only one possibility for the dimension of S (namely, $\dim(S\backslash q\backslash r)+rk^*(\rho)$) so that S must be $(S\backslash q)+(S\backslash r)$. The converse is clear.) If ρ is $U_0^2(q,r)$ or $U_0^1(q)+U_1^1(r)$ or $U_1^1(q)+U_0^1(r)$ then ρ is uniquely determined by $\rho\backslash q$ and $\rho\backslash r$. But if ρ is $U_1^2(q,r)$ or $U_2^2(q,r)$ then in both these cases $\rho\backslash q=U_1^1(r)$ and $\rho\backslash r=U_1^1(q)$, and in neither case is ρ uniquely determined by $\rho\backslash q$ and $\rho\backslash r$. This corresponds to the case when $\dim(S\backslash q\backslash r)=\dim(S\backslash q)=\dim(S\backslash r)$, so that $S\backslash q\backslash r=S\backslash q=S\backslash r$ which therefore equals $(S\backslash q)+(S\backslash r)$ (as subspaces of R^Q). In this case, either $\dim(S)=\dim(S\backslash q\backslash r)$ (corresponding to $\rho=U_2^2(q,r)$) so that $S=S\backslash q\backslash r$, or else $\dim(S)=\dim(S\backslash q\backslash r)+1$ (corresponding to $\rho=U_1^2(q,r)$) so that $S\backslash q\backslash r$ is a proper subspace of S and S has as a basis, that of $S\backslash q\backslash r$ together with some vector $a \in R^Q$ such that $\{q,r\} \subseteq \mathrm{supp}(a)$. Also in this case either $\alpha(S)\backslash q\backslash r=\rho=U_2^2(q,r)$ so that q and r are coloops of $\alpha(S)$, or else $\alpha(S)\backslash q\backslash r=\rho=U_1^2(q,r)$, so that q and r are coparallel in $\alpha(S)$. So σ is uniquely determined by $S\backslash q$ and $S\backslash r$ unless q and r are coloops of $\alpha(S)$ or are coparallel in $\alpha(S)$. By duality, S is uniquely determined by S/q and S/r unless q and r are loops of $\alpha(S)$ or are parallel in $\alpha(S)$.

We now find the ψ–structures of $\mathscr{D}(R)$. The minor class $\mathscr{D}(R)$ has a unique 0–point structure ($R^\emptyset$ is the only subspace of $R^\emptyset$) so that the two structures $\mathrm{loop}(q)=U_0^1(q)$ and $\mathrm{coloop}(q)=U_1^1(q)$, on ground set $\{q\}$, are ψ–equivalent. Suppose that $S \in \mathscr{D}(R)$ on ground set Q, is another ψ–structure, so that the order $|Q|$ of S is at least 2. Now S is not uniquely determined by $S\backslash q$ and $S\backslash r$, nor by S/q and S/r, for any distinct $q,r \in Q$. Since no loop can be a coloop and no loop or coloop can be parallel or coparallel to any other element it follows that all elements of Q are both parallel and coparallel in $\alpha(S)$ to all other elements of Q. Therefore $\alpha(S)$ is $U_1^2(q,r)$ (for some 2–element ground set $\{q,r\}$) (since all 1–element subsets of Q have rank and corank 1 but no 2–element subset of Q have rank or corank 2, so that $\mathrm{rk}(\alpha(S))=\mathrm{rk}^*(\alpha(S))=1$ and hence $|Q|=1+1=2$). Also, all $S \in \mathscr{D}(R)$ with $\alpha(S)=U_1^2(q,r)$ are ψ–equivalent to each other, since they all have the same 1–point minors. Now $\dim(S)=\mathrm{rk}^*(\alpha(S))=2-1=1$. Write the elements of $R^{\{q,r\}}$ as ordered pairs with the first and second component, labelled by q and r respectively, as before. Therefore S is a one dimensional subspace of $R^{\{q,r\}}$ and since $S\backslash q=\mathrm{loop}(r)$ and $S\backslash r=\mathrm{loop}(q)$, so that $R^{\{q\}} \cap S=R^{\{r\}} \cap S=\{0\}$, it follows that S is of the form $\{(c,ca) \mid c \in R\}$ for some non–zero $a \in R$. Call this structure (a,q,r) the *slope* with slope a *of* r *over* q. Observe that (a,q,r) is equal to (a^{-1},r,q). So (a,q,r) is isomorphic and ψ–equivalent to (a^{-1},q,r) (but they are unequal unless a=1 or a=–1). So the ψ–structures on ground set $\{q,r\}$ are the slopes (a,q,r) for all non–zero $a \in R$, and with loop(q) and coloop(q), these are the only ψ–structures.

Actually, when $|R|=2$, there is only one non-zero $a \in R$, and hence only one slope (a,q,r), so that in this case there are no 2-point ψ-structures.

Any structure $S \in \mathscr{D}(R)$, on ground set Q, is uniquely determined by specifying all its 1-point minors, and for those distinct $q,r \in Q$ and $P \subseteq Q-\{q,r\}$ such that $(\alpha(S)/P)\cdot\{q,r\}=\alpha((S/P)\cdot\{q,r\})=U_1^2(q,r)$, specifying for which non-zero $a \in R$ it holds that $(S/P)\cdot\{q,r\}$ is (a,q,r). Visualising this as a patterned Q-hypercube, as described in section 11, all the 1-faces (edges) are labelled as loop or coloop, exactly as for the matroid $\alpha(S)$, and all the (square) 1-faces which look like 2e (that is U_1^2) are labelled to indicate which slope the corresponding 2-point minor is. (All the other 2-faces, as well as the m-faces for $m=0,3,4,5,\cdots,|Q|$ are left blank.) The patterned Q-hypercube depicting $\alpha(S)$ is obtained from that depicting S, simply by "blanking" the above patterned 2-faces and leaving the 1-faces unchanged. Also the dual of (a,q,r) is $(-a^{-1},q,r)$ so that the patterned Q-hypercube depicting S^* is obtained from that depicting S, by sending each point $(x_q|q \in Q)$ of the Q-hypercube to $(1-x_q|q \in Q)$ (since deletion and contraction are swapped as for the matroid case in section 12), changing each 1-face labelled loop to coloop and visa-versa, and changing each slope $a \in R-\{0\}$ to $-a^{-1}$.

The other subspaces of $R^{\{q,r\}}$ are $R^{\{q,r\}}$, $R^{\{q\}}$, $R^{\{r\}}$ and $R^{\emptyset}$ and α sends these to $U_0^2(q,r)$, $U_0^1(q)+U_1^1(r)$, $U_1^1(q)+U_0^1(r)$ and $U_2^2(q,r)$ respectively, which are all the remaining matroids on $\{q,r\}$. Naming the two 1-point structures in $\mathscr{D}(R)$ as loop and coloop, as we have, it follows that $\mathscr{D}(R)$ has the same six 2-point natural excluded minors as $\mathscr{M}$, namely 2b,2c,2d,2g,2h and 2i (see figure 1).

Now the homomorphism $\alpha:\overline{\mathscr{D}(R)}\to\mathscr{M}$ sends all (and only those) elements of $\overline{\mathscr{D}(R)}$exc(2b,2c,2d,2g,2h,2i) into $\mathscr{M}$. In particular, α sends any other natural excluded minor (which must have order at least 3) of $\mathscr{D}(R)$ into $\mathscr{M}$.

Suppose that $S \in \overline{\mathscr{D}(R)}$, with ground set Q, and of order $|Q| \geq 3$, is a natural excluded minor of $\mathscr{D}(R)$. Suppose also that the matroid $\alpha(S)$ has a coloop $q \in Q$. So $\alpha(S\backslash q)=(\alpha(S)\backslash q)=(\alpha(S)/q)=\alpha(S/q)$ but $(S\backslash q)\neq(S/q)$ as shown below. (Observe that all the minors of S which are ψ-structures, are covered by specifying $S\backslash q$, S/q and the fact that q is a coloop of $\alpha(S)$, so that these uniquely determine S.) Now if $S\backslash q$ were equal to S/q then these can be treated as subspaces of R^Q, and S must be $S\backslash q$. But then $S \in \mathscr{D}(R)$ contradicting the assumption that it was a natural excluded minor of $\mathscr{D}(R)$. Therefore $(S\backslash q)\neq(S/q)$, as claimed. Since $S\backslash q$ and S/q agree on their 1–point minors, they must disagree on the slope of a corresponding 2–point minor. That is, for some distinct $p,r \in Q-\{q\}$ and some $P \subseteq Q-\{p,q,r\}$ it holds that $(S\backslash q/P)\cdot\{p,r\}=(a,p,r)$ and $(S/q/P)\cdot\{p,r\}=(b,p,r)$ where a and b are distinct non–zero elements of R, and $(S/P)\cdot\{p,r\}$ is not in $\mathscr{D}(R)$. By the minimality of S it follows that $Q=\{p,q,r\}$ and S is uniquely specified by saying that $\alpha(S)=U_1^1(q)+U_1^2(p,r)$ and $S\backslash q=(a,p,r)$ and $S/q=(b,q,r)$ for some distinct non–zero $a,b \in R$. All the natural excluded minors S, of $\mathscr{D}(R)$, where $\alpha(S)$ is a matroid with a coloop, are given by considering all pairs of distinct non–zero $a,b \in R$. By duality any natural excluded minor S, of $\mathscr{D}(R)$, where $\alpha(S)$ is a matroid with a loop, are given by $\alpha(S)=U_0^1(q)+U_1^2(q,r)$ and $S\backslash q=(a,q,r)$ and $S/q=(b,q,r)$ for all distinct non–zero $a,b \in R$. (Recall that the natural excluded minors are obtained from the natural excluded isominors, by partitioning the latter according to isomorphism and choosing one from each partition. But above, all the natural excluded isominors (of a certain form) on a particular ground set are given, so that some are mentioned more than once (at most twice in this case, and at most 4!=24 time in a later case) up to isomorphism. Obviously this is not a problem, and will not attract further comment.)

Suppose now that $S \in \overline{\mathscr{D}(R)}$, with ground set Q, is a natural excluded minor of $\mathscr{D}(R)$, and that $\alpha(S)$ is a matroid with no loops or coloops, but with distinct $q,r \in Q$ being parallel. Observe that if S is given by $\alpha(S)=U_1^3(p,q,r)$

and the slopes S\p,S\q and S\r are respectively (a,q,r), (b,r,p) and (c,p,q) where non–zero a,b,c∈R are such that abc≠–1, then S is a natural excluded minor of $\mathscr{D}$(R), as is easily verified. Assume instead, that S is not isomorphic to a structure of this form. Now α(S·{q,r})=α(S)·{q,r}=U_1^2(q,r) (by the definition of parallel) so that S·{q,r} is the slope (a,q,r) for some non–zero a∈R. If P⊆Q–{q,r} is such that α((S/P)·{q,r})=(α(S)/P))·{q,r}=U_1^2(q,r), then by the properties of the partial order ≺ with respect to deletions and contractions, it follows that if p∈P then (α(S/P–{p})·{q,r})=(α(S)/P–{p})·{q,r}=U_1^2(q,r). By the exclusion of the natural excluded minors mentioned in the previous paragraph, it follows by induction that (S/P)·{q,r}=(a,q,r) for every P⊆Q–{q,r} such that α(S/P)·{q,r}=U_1^2(q,r). Since q is a loop of α(S)/r=α(S/r) and r is a loop of α(S)/q=α(S/q) and by the exclusion of the natural excluded minors mentioned in earlier paragraphs it follows that S/q\r=S/q/r=S/r/q=S/r\q. Of course S\q\r=S\r\q. As shown earlier, α(S) has the bijection from Q to Q, which fixes Q–{q,r} and swaps q and r, as an automorphism. So the bijection from Q–{r} to Q–{q}, which fixes Q–{q,r} and sends q to r, is an isomorphism from α(S)\r to α(S)\q. So for any p∈Q–{q,r} and any P⊆Q–{p,q,r} it follows that (α(S)\r/P)·{p,q} (which equals (α(S)/P)·{p,q}) is U_1^2(p,q) if and only if (α(S)\q/P)·{p,r} (which equals (α(S)/P)·{p,r}) is U_1^2(p,r). For any p and P such that these equivalent statements are true, the only possibility for α((S/P)·{p,q,r})=(α(S)/P)·{p,q,r} is U_1^3(p,q,r), by the exclusion of 2b,2h,2c and 2d (see figure 1). (The dual of this statement was shown in the paragraph defining coparallel.) In this case (S/P)·{p,q} is (b,p,q) for some non–zero b∈R, and by the exclusion of the natural excluded minors described in this paragraph, it follows that (S/P)·{p,r} is (–1/ab,r,p) which is (–ab,p,r). All the minors of S which are ψ–structures have been specified, and in the absence of the abovementioned natural excluded minors, they are uniquely determined by S\r and S·{q,r}=(a,q,r). Interpreting elements (d_t|t∈Q–{r})∈(S\r)⊆$R^{Q-\{r\}}$

as $(d_t|t\in Q)\in R^Q$ with $d_r=0$, and letting $(e_t|t\in Q)$ be such that $e_q=1$, $e_r=a$, and $e_t=0$ for every $t\in Q-\{q,r\}$, then this unique S is actually $\{(d_t|t\in Q)+f(e_t|t\in Q)\,|\,(d_t|t\in Q)\in S\backslash r \text{ and } f\in R\}$ which is a subspace of R^Q. This contradicts the assumption that S is not in $\mathscr{D}(R)$, so there is no such S. Hence the only natural excluded minors S, on ground set Q, of $\mathscr{D}(R)$ such that $\alpha(S)$ is a matroid with no loops or coloops, but with some distinct $q,r\in Q$ being parallel are those mentioned earlier in the paragraph. By duality the only possibilities for S, with parallel changed to coparallel (in the previous sentence) are (up to isomorphism) when S is given by $\alpha(S)=U_2^3(p,q,r)$ and the slopes $S/p, S/q, S/r$ are respectively $(a,q,r),(b,r,p),(c,p,q)$ where non-zero $a,b,c\in R$ are such that $(-1/a)(-1/b)(-1/c)\neq-1$, that is, $abc\neq1$.

Suppose that $S\in\overline{\mathscr{D}(R)}$, with ground set Q, is a natural excluded minor of $\mathscr{D}(R)$ which is not isomorphic to any of those mentioned above. So $\alpha(S)$ is a matroid with no loops or coloops and no pairs of elements being parallel or coparallel, and hence every 2-element subset of Q has rank and corank 2 and $\alpha(S)$ has rank and corank at least 2. Suppose that the corank of $\alpha(S)$ is at least 3 so that there exists $\{1,2,3\}\subseteq Q$ with corank 3 and $\alpha(S)/(Q-\{1,2,3\})=U_0^3(1,2,3)$. Let $T\in\mathscr{D}(R)$, with ground set Q, be $(S\backslash1)+(S\backslash2)$ (as defined earlier) so that $S\backslash1=T\backslash1$ and $S\backslash2=T\backslash2$. Now $S\backslash3$ is uniquely determined by $S\backslash3\backslash1$ and $S\backslash3\backslash2$, since $\alpha(S)\backslash3/Q-\{1,2,3\}=U_0^3(1,2,3)\backslash3=U_0^2(1,2)$. But $S\backslash3\backslash1=S\backslash1\backslash3=T\backslash1\backslash3=T\backslash3\backslash1$ and similarly $S\backslash3\backslash2=T\backslash3\backslash2$ and these uniquely determine $T\backslash3$ which therefore equals $S\backslash3$. Also $S/3$ is uniquely determined by $S/3\backslash1$ and $S/3\backslash2$ since $\alpha(S)/3/Q-\{1,2,3\}=U_0^3(1,2,3)/3=U_0^2(1,2)$. But $S/3\backslash1=S\backslash1/3=T\backslash1/3=T/3\backslash1$ and similarly $S/3\backslash2=S/3\backslash1$ and these uniquely determine $T/3$ which therefore equals $S/3$. Similarly if $q\in Q$, then since $\{1,2,3\}-\{q\}$ has at least the two elements needed in the argument, $S/q=T/q$. Also if $q\in Q-\{1,2,3\}$ then $S\backslash q$ is uniquely determined by $S\backslash q/1$ and $S\backslash q/2$, since $\alpha(S)\backslash q\backslash Q-\{1,2,q\}=\alpha(S)\backslash Q-\{1,2\}$ is not $U_0^2(1,2)$ or $U_1^2(1,2)$, by the assumption that $\alpha(S)$ has no loops or parallel pairs of elements. But $S\backslash q/1=S/1\backslash q=T/1\backslash q=T\backslash q/1$ and similarly $S\backslash q/2=T\backslash q/2$ and these uniquely determine $T\backslash q$ which thus equals $S\backslash q$. Therefore $S\backslash q=T\backslash q$ and $S/q=T/q$ for all $q\in Q$ and hence $S\psi T$, contradicting the assumption that $S\notin\mathscr{D}(R)$, while $S\in\mathscr{D}(R)$. So the assumption that $rk^*(\alpha(S))\geq 3$ is false, so that $rk^*(\alpha(S))=2$. By duality $rk(\alpha(S))=2$ as well, and the only possibility for $\alpha(S)$ is $U_2^4(1,2,3,4)$ for some 4-element ground set $\{1,2,3,4\}$. Now U_1^2 is obtained as a minor of U_2^4 by deleting any element and contracting any other. For distinct $k,\ell,m,n\in\{1,2,3,4\}$ let non-zero $a_{mn}^{k\ell}\in R$ be such that $S\backslash k/\ell=(a_{mn}^{k\ell},m,n)$. (Note that $a_{mn}^{k\ell}=1/a_{nm}^{k\ell}$.) Let $a=a_{23}^{14}$, $b=a_{31}^{24}$, $c=a_{12}^{34}$, $d=a_{43}^{21}$, $e=a_{24}^{31}$, $f=a_{32}^{41}$, $g=a_{41}^{32}$, $h=a_{13}^{42}$, $i=a_{34}^{12}$, $j=a_{21}^{43}$, $k=a_{42}^{13}$, $\ell=a_{14}^{23}$. By the exclusion of the natural excluded minors mentioned in the previous paragraph, it follows that $1=abc=def=ghi=jk\ell$ and $-1=aik=bd\ell=ceg=fhj$. Of these eight equations, no six determine the others, but any seven determine the eighth. (So when $|R|$ is finite there are $(|R|-1)^{12-7}=(|R|-1)^5$ possibilities). Now any $S\in\mathscr{D}(R)$ with $\alpha(S)=U_2^4(1,2,3,4)$ is of the form $\{m(1,0,w,x)+n(0,1,y,z)\,|\,m,n\in R\}$ where non-zero $w,x,y,z\in R$ with $wz\neq xy$. (When $|R|$ is finite there are $(|R|-1)^3(|R|-2)$ possibilities, so there are clearly many natural excluded minors.) It is routine to check that in this case $a=(yx-wz)/x$, $b=z/(wz-xy)$, $c=-x/z$, $d=y/z$, $e=z$, $f=1/y$, $g=1/x$, $h=w$, $i=x/w$, $j=-y/w$, $k = w/(wz-xy)$ and $\ell=(xy-wz)/y$ and, for example, $afi=i-ef$. (Similar equations to this obtained by the symmetry of the situation are also obtained from the equation in the presence of the eight given above.) Conversely the equation $afi=i-ef$ is sufficient to ensure that S is in $\mathscr{D}(R)$ (simply put $x=1/g$, $y=1/f$, $w=h$, $z=e$). So S is a natural excluded minor of $\mathscr{D}(R)$ with $\alpha(S)=U_2^4(1,2,3,4)$ if and only if the eight equations above hold but $afi\neq i-ef$. This exhausts all the possibilities for natural excluded minors.

So in summary, the ψ-structures of $\mathscr{D}(R)$ are loop(q) and coloop(q) for a

1–element ground set {q} and the slopes (a,q,r) for all non–zero $a \in R$ for a 2–element ground set {q,r} (unless $|R|=2$). The natural excluded minors of $\mathscr{D}(R)$ have order 2 (these are the same six as for $\mathscr{M}$) and 3 (some structures in $\mathscr{D}(R)$ which α sends to $U_0^1+U_1^2$, $U_1^1+U_1^2$, U_1^3 and U_2^3) and 4 (some structures in $\overline{\mathscr{D}(R)}$ which α sends to U_2^4). This holds for any field R, finite or infinite.

The Minor Class $\mathscr{M}(R)$

For any ring R let $\mathscr{M}^C(R)$ denote the homomorphic image $\alpha(\mathscr{D}(R))$ of $\mathscr{D}(R)$ under α, and if R is a division ring denote it as $\mathscr{M}(R)$. The elements of $\mathscr{M}(R)$ are *the matroids coordinatisable over* R. Matroid theorists are particularly interested in knowing the excluded minors of $\mathscr{M}(R)$ in $\mathscr{M}$. (This immediately yields the ψ–description of $\mathscr{M}(R)$, and visa versa as shown in section 12.) A conjecture of Rota [12] states that when R is a finite field, $\mathscr{M}(R)$ has finitely many excluded minors in $\mathscr{M}$ (or equivalently, $\mathscr{M}(R)$ has finitely many natural excluded minors). It is not surprising that $\mathscr{M}(R)$ generally has infinitely many natural excluded minors when R is infinite, and this seems compatible with the fact that, while the order of the ψ–structures and natural excluded minors of $\mathscr{D}(R)$ is bounded, there are infinitely many of them. But when R is finite, $\mathscr{D}(R)$ has finitely many ψ–structures and natural excluded minors, and it is not unreasonable to believe that $\mathscr{M}(R)$ also has finitely many ψ–structures (which it does) and natural excluded minors (the subject of much research).

Consider the visualisation of structures in $\mathscr{D}(R)$. They are patterned hypercubes with patterns on the 1–faces and some of the 2–faces, as described earlier. In fact they are exactly those in which certain patterned 2–,3– and 4–hypercubes do not appear as patterned subhypercubes. The surjective homomorphism to $\mathscr{M}(R)$ simply erases the pattern on the 2–faces, and it does not seem unreasonable to ask now which patterned hypercubes minimally do not appear as patterned subhypercubes. However this is not a simple problem and it is probably unrealistic to expect an answer for general finite fields R. But it would seem to be feasible that a theorem about minor class homomorphisms with certain conditions (or even homomorphisms of algebras with certain conditions) could yield as a corollary that $\mathscr{M}$(R) has finitely many natural excluded minors. On the one hand, the latter is a monumental problem which, despite great effort, has not been solved. On the other hand we now have a large body of universal algebra at our disposal; of course there are many unsolved problems in algebra.

Let us consider $\mathscr{D}$(R) and $\mathscr{M}$(R) when $|R|=2,3$ or 4, the cases which have appeared in the literature. Since any finite field R is uniquely determined up to isomorphism by its cardinality $|R|$, then $\mathscr{D}$(R) and $\mathscr{M}$(R) can be unambiguously denoted $\mathscr{D}(|R|)$ and $\mathscr{M}(|R|)$. When $|R|=2$, there are, as noted earlier, no 2–point ψ–structures in $\mathscr{D}(2)$. So the homomorphism, α, from $\mathscr{D}(2)$ to $\mathscr{M}(2)$ is actually an isomorphism. Now $\mathscr{D}(2)$ has the usual six 2–point natural excluded minors, but since there do not exist distinct non–zero $a,b\in R$ nor non–zero $a,b,c\in R$ with $abc\neq 1=-1$, there are no 3–point natural excluded minors. On any four element ground set there is $(|R|-1)^5=1$ possible structure which α sends to U_2^4 in $\mathscr{M}$ of which all but $(|R|-1)^3(|R|-2)=0$ are natural excluded minors. (That is, $\mathscr{D}(2)$ has one natural excluded minor of order four, which α sends to U_2^4.) This provides an independent proof of the well known theorem of Tutte [14] that $\mathscr{M}(2)$ has only U_2^4 as an excluded minor in $\mathscr{M}$.

Now $\mathscr{D}(3)$ has (up to isomorphism) six 2–point, eight 3–point and two 4–point natural excluded minors, but $\mathscr{M}(3)$ has ([R. Reid, unpublished, 1970]) [2,13] six 2–point, (those of $\mathscr{M}$) two 5–point (U_2^5 and its dual U_3^5) and two 7–point (the so called Fano matroid [18] and its dual) natural excluded minors.

And $\mathscr{D}(4)$ has six 2–point, twelve 3–point and seventeen 4–point natural excluded minors, whereas $\mathscr{M}(4)$ has at least six 2–point (those of $\mathscr{M}$) three 6–point (including U_2^6 and its dual U_4^6), two 7–point, and one 8–point natural excluded minors [10], but it is not known if there are others.

These examples demonstrate that the natural excluded minors of a homomorphic image do not bear an obvious relation to those of its source. The examples in the next section further emphasise this.

The ψ–description $\mathscr{D}(\mathrm{R})$ might not be too hard to find for certain kinds of rings R more general than fields. In particular, if R is a principal ideal domain, or one of a certain unknown class of non–commutative rings (including non–commutative division rings) then there is still a sensible notion of dimension of modules over R. There is no duality in the non–field case, but some of the arguments used in the field case can be salvaged. The ring of integers $\mathbb{Z}$ is a principal ideal domain and the ψ–description of $\mathscr{D}(\mathbb{Z})$ is probably relatively simple. Many interesting sub minor classes of $\mathscr{M}$ and $\mathscr{M}^{\mathrm{C}}$ are homomorphic images of $\mathscr{D}(\mathbb{Z})$ or of its sub minor classes. For example $\mathscr{M}^{\mathrm{C}}(\mathbb{Z})$ and $\mathscr{M}(\mathbb{Q})$ as well as the minor class of *regular matroids* (see [15]). In general, if R is an integral domain and Q(R) is its quotient field (so that $\mathbb{Q}=\mathrm{Q}(\mathbb{Z})$), then $\mathscr{M}(\mathrm{Q}(\mathrm{R}))$ is the homomorphic image of $\mathscr{D}(\mathrm{R})$ via the homomorphism which, for each 1–element ground set {q}, sends $\emptyset^{\{q\}}$ to coloop(q) and all other structures in $\mathscr{D}(\mathrm{R})$ on ground set {q}, to loop(q).

SECTION 14: MINOR CLASSES OF GRAPHS

A *graph* G [16] consists of a finite set V(G) whose elements are called *vertices,* a finite set E(G) whose elements are called *edges,* and a relation of *incidence*, which associates with each edge, two vertices (not necessarily distinct), called its *ends.* Two edges are *adjacent* if they are incident with the same vertex (that is, an end of one coincides with an end of the other) and two vertices are *adjacent* if they are the two ends of one edge. An edge is a *loop* if its two ends are equal, and a *link* if its two ends are distinct. Two edges are *parallel* if they are both links and they have the same ends. A *directed graph* or *digraph* G, like a graph, consists of a finite set of vertices V(G) and a finite set of edges E(G), but now the two ends of each edge distinguished (giving each edge a "direction") so that for any edge, the vertex at one end is called its *head* and the vertex of the other end is called its *tail.* (Again these need not be distinct.) Obviously a graph can be obtained from a digraph by "ignoring the direction of each edge", that is, letting the head and tail of an edge be simply its two (undistinguished) ends. A graph or digraph is *simple* if it has no loops or parallel edges.

A graph can be depicted by drawing its vertices as points, and each edge as a smooth curve between the points representing its ends (and touching no other points). A digraph can be depicted in the same way except that an arrow is placed on each edge pointing from tail to head.

An *isomorphism* from graph G to graph H, consists of a bijection from V(G) to V(H) and a bijection from E(G) to E(H) which preserve incidence. If such an isomorphism exists, then G and H are *isomorphic.* If both the bijections can be chosen to be the identity function then G and H are *equal.* We shall be constructing minor classes of graphs in which the ground sets are either edge sets or vertex sets, but not both. When the ground sets are vertex sets, it is immaterial how the edges are named. Two graphs are *vertex–equal* if there is an isomorphism between them for which the bijection between the vertex sets is the identity function, regardless of the bijection between the edge sets. *Vertex–graphs* (this term has a different meaning in [16]) or *graphs with anonymous edges*, are graphs for which isomorphism is defined as usual, but equality is defined to be vertex–equality. An *isolated vertex* is a vertex incident with no edges. When ground sets are edge sets, it is immaterial how the vertices are named. Two graphs, with no isolated vertices, are *edge–equal* if there is an isomorphism between them for which the bijection between the edge sets is the identity function, regardless of the bijection between vertex sets. *Edge graphs* or *graphs with anonymous vertices* are graphs with no isolated vertices, with the usual definition of isomorphism, but with equality defined to be edge–equality. *Vertex–digraphs* and *edge–digraphs* are defined similarly.

To save repetition, it will be now said once, that all the alleged minor classes below are indeed minor classes. (They are all standard in the literature, and those properties which establish them as minor classes are well known.) The relevant version of structure isomorphism is as defined above for each type of graph.

The minor class $\mathscr{G}$(VG) has vertex–graphs as its structures, finite sets as its ground sets (vertex sets), and one manner of point removal called *vertex deletion.* For any vertex–graph G$\in\mathscr{G}$(VG), and any vertex v in its ground set V(G), the deletion of v from G is denoted G\v, and is obtained from G by removing v from V(G) and removing all edges incident with v. There is a unique 0–vertex graph (with no vertices and hence no edges) and so all the 1–vertex graphs (with a single vertex and n loops, for any non–negative integer n) are ψ–equivalent. Any two 2–vertex graphs, with 2–element vertex set {v,w} say, are ψ–equivalent if they have the same number of loops incident with v, and the same number incident with w. They can have any non-negative integer number of edges (links) incident with both v and w. These ψ-structures specify the number of edges incident with any pair of (not necessarily distinct) vertices and conversely since these numbers can be chosen arbitrarily, the ψ-structures can be chosen arbitrarily as minors of a graph. So the ψ-structures mentioned are the only ones and there are no natural excluded minors, so that $\mathscr{G}$(VG) is a complete minor class. It is actually isomorphic to the direct product of two of its sub minor class (which are also complete) namely that containing all graphs with only loops as edges (this minor class has 1-vertex ψ-structures), and that containing all graphs with only links as edges (this minor class has 2-vertex ψ-structures). The sub minor class $\mathscr{G}$(SVG), of $\mathscr{G}$(VG), containing only the simple vertex-graphs has only two ψ-structures, namely a 2-vertex graph either with or without an edge incident with both vertices. Specifying all 2-point minors of a simple graph determines for each pair of distinct vertices, whether or not there is an edge between them (that is, whether or not they are adjacent) thus determining the graph (so there are indeed only two ψ-structures). These can clearly be specified arbitrarily, so that there are no natural excluded minors and $\mathscr{G}$(SVG) is complete.

Similar definitions and comments apply to $\mathscr{G}$(VD) the minor class of vertex-digraphs with one manner of point removal, namely vertex deletion. A vertex is deleted from a digraph, exactly as it is from a graph. This time the 2-element ψ-structures specify how many links there are in each direction between any two vertices. Again $\mathscr{G}$(VD) is complete and is isomorphic to the direct product of two of its complete sub minor classes. Also $\mathscr{G}$(SVD), containing only the simple vertex-digraphs, is a complete sub minor class of $\mathscr{G}$(VD). It has three 2-vertex ψ-structures (these are the only ones) since a pair of vertices can have no edge between them or a directed-edge in either direction. The latter two are isomorphic but not equal (as ψ-structures sometimes are). Observe that there are surjective homomorphisms from

$\mathscr{G}$(VD) to $\mathscr{G}$(VG) and from $\mathscr{D}$(SVD) to $\mathscr{G}$(SVG) which send each vertex–digraph to the corresponding vertex–graph, by removing the direction of each edge.

The minor class $\mathscr{G}$(ED,1) has edge–digraphs as its structures, finite sets as its ground sets (edge sets), and one manner of point removal called *edge deletion.* For any edge–digraph G$\in\mathscr{G}$(ED,1) and any edge e in its ground set E(G), the *deletion* of e from G is denoted G\e, and is obtained from G by removing e from E(G) and removing any isolated vertices this may have created. For any P$\subseteq$E(G), the graph G\P is obtained from G by deleting every edge in P (in any order) and G$\cdot$P denotes G\(E(G)–P). Now an edge–digraph may be described in terms of an equivalence relation $\sim$ on E(G)$\times${h,t} for some two element set {h,t}. For edges e,f$\in$E(G) let (e,h)$\sim$(f,t) if and only if the head of edge e in G is the same vertex as the tail of edge f in G, and define this relation on the rest of E(G)$\times${h,t} in the obvious way. Clearly the relation $\sim$, uniquely determines G, and also, any equivalence on Q$\times${h,t} defines (uniquely) an edge–digraph with edge set Q. For any edge–digraph G, the relation $\sim$ is determined by the 1– and 2–edge minors of G, since these determine whether or not the equivalence $\sim$ holds between (e,x) and (f,y) for every e,f$\in$E(G) and every x,y$\in${h,t}. (Note that |{e,f}| is either 1 or 2.) Since there is a unique 0–edge digraph (with no edges, and hence no vertices) it follows that the ψ–structures have order 1 and 2. In searching for natural excluded minors, we specify the 1– and 2–edge minors of a structure G, with ground set Q say, in the completion $\overline{\mathscr{G}(\mathrm{ED},1)}$, of $\mathscr{G}$(ED,1). This will give a well defined relation $\sim$ on Q$\times${h,t} which is reflexive and symmetric, so that the only way G will not be in $\mathscr{G}$(ED,1) is if $\sim$ is not transitive. In this case there exist e,f,g$\in$Q and x,y,z$\in${H,t} such that (e,x)$\sim$(f,y) and (f,y)$\sim$(g,z) but (e,x)$\sim$(g,z) does not hold, so that G$\cdot${e,f,g} is also not in $\mathscr{G}$(ED,1). Therefore the natural excluded minors of $\mathscr{G}$(ED,1) have order of most 3, and it is easily checked that they do not have order less than 3. The ψ-structures are illustrated in figure 2, with the convention that ψ-equivalent graphs are illustrated in the same row. (The reason for the dotted line in row 4 becomes apparent later.) The ground set elements, that is edges, have not been labelled to avoid cluttering the diagram, but they should be, since it is necessary to specify an actual edge graph. So a convention is adopted where all the graphs in the same row have the same edge set, and for all the 2-edge graphs in the same row, the upper edges are labelled by one element of the ground set, while the lower edges are labelled by the other. Observe that the graphs, second and third from the right in row 4 of figure 2, are isomorphic but not equal

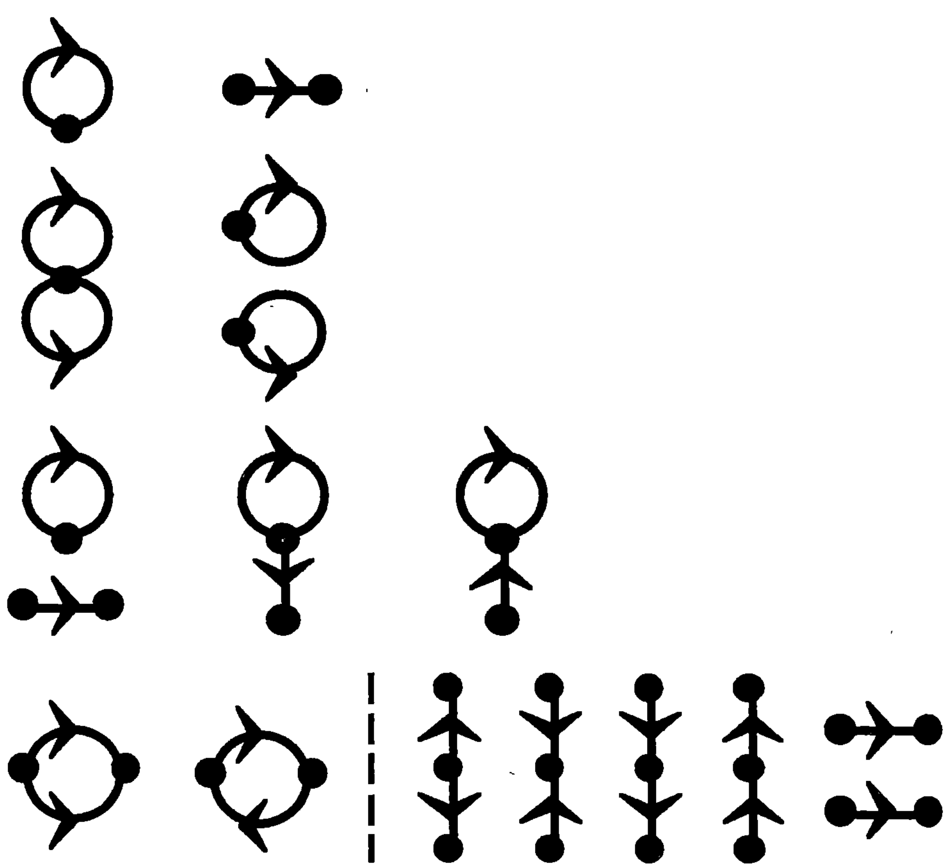

Figure 2: The ψ-structures of $\mathscr{G}$(ED,1).

There are too many natural excluded minors to list, but at least we can count them (and readers can try to obtain the same result). There are two possible criteria for counting them. The correct way is to count them up to isomorphism, using either Polya enumeration, or simply counting by observation. Alternatively, one can count all the natural excluded isominors on a fixed (3–element, in this case) ground set, which is sometimes easier. The latter method gives a result at least as large, since for each structure counted, so are all structures isomorphic to it, on the same ground set. First consider structures in $\overline{\mathscr{G}(\mathrm{ED},1)}$ on a particular 3–element ground set, in which all three 1–point minors are loops. There are 2^3 possibilities for the 2–point minors (see figure 2, row 2) making 8 possible structures, or 4 up to isomorphism (since either 0 or 1 or 2 or 3 of the three 2–point minors could be the left hand graph in row 2 of figure 2), and we denote this as 4:8 possible structures. (The first figure counts up to isomorphism, the second counts on a fixed ground set.) Of these, 3:5 are actually graphs, so that 4:8–3:5=1:3 are natural excluded minors. Similarly, when two (respectively one, none) of the three 1–point minors are loops there are 12:54–6:24=6:30 (respectively 39:189–17:78=22:111, 73:343–24:84=49:259) natural excluded minors. So there are a total of 1+6+22+49=78 natural excluded minors (or 3+30+111+259=403 on a fixed ground set).

The minor class $\mathscr{G}(\mathrm{EG},1)$ has edge–graphs as its structures, finite sets as its ground sets (edge sets), and one manner of point removal called *edge deletion*, which is defined and denoted exactly as it is for $\mathscr{G}(\mathrm{ED},1)$. So there is a surjective homomorphism from $\mathscr{G}(\mathrm{ED},1)$ to $\mathscr{G}(\mathrm{EG},1)$ which sends each edge–digraph to the corresponding edge–graph, by removing the direction of each edge. Figure 3 illustrates some ψ–structures of $\mathscr{G}(\mathrm{EG},1)$. Graphs in the same row are ψ–equivalent, and the same convention is used for labelling edges, as with figure 2. Also the 3–edge graphs in row 5 should be labelled by the same three elements, but due to their symmetry this can be done arbitrarily

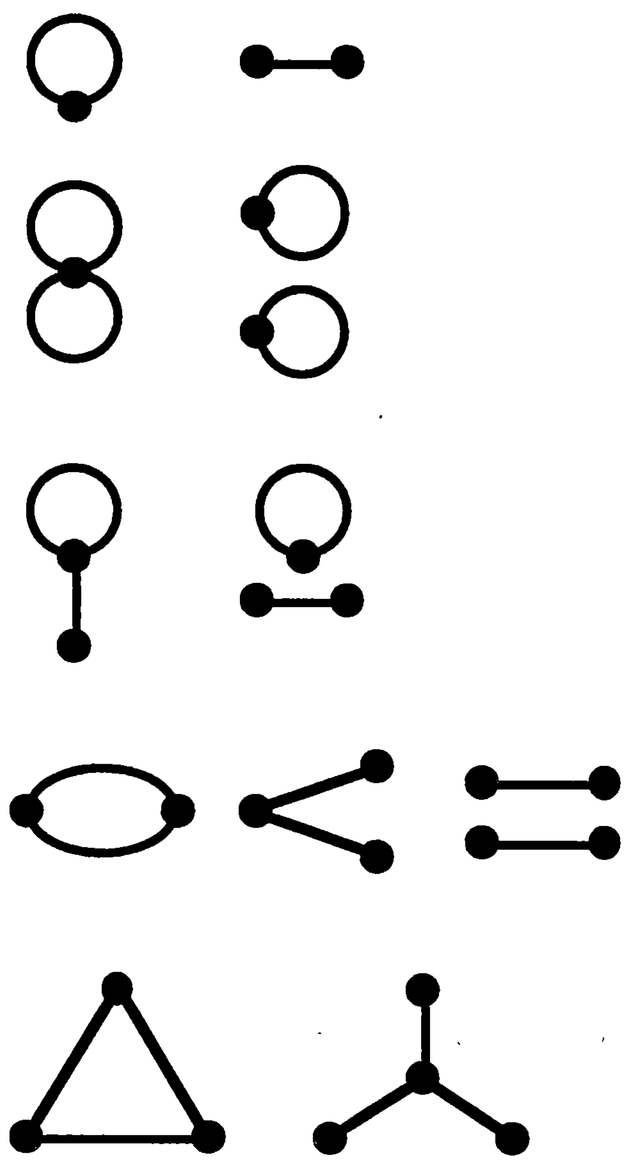

Figure 3: The ψ–structures of $\mathscr{G}$(EG,1).

It is routine to show that all the edge–graphs in any row in figure 3, are ψ–equivalent, and that all the ψ–structures of order at most 3, are illustrated. Suppose edge–graph G, with ground set (edge set) Q, is a ψ–structure in $\mathscr{G}$(EG,1) with order minimally greater than 3. Suppose all the minors of G which appear in figure 3 are specified, which uniquely determines all proper minors of G, by assumption. In particular, choosing some edge $e \in Q$, it follows that $G \backslash e$ is uniquely determined. We shall examine how the edge e can be attached to $G \backslash e$, consistent with the given proper minors of G. The 1–edge minor $G \cdot e$ determines whether e is a loop or a link and for edge $f \in Q-\{e\}$, the 2–edge minor $G \cdot \{e,f\}$ determines the way in which the ends of e and f coincide (if at all). If e shares no end in G (that is, is not adjacent) with any other edge then this is determined by the 2–edge minors of G, and G is constructed from $G \backslash e$ by adding edge e with its ends being new vertices added for the purpose. If edge e is parallel to edge f in G, then $G \cdot \{e,f\}$ shows this, and G is constructed from $G \backslash e$ by adding edge e with its ends being the ends of f. Without loss of generality, assume that G has no two edges parallel, but that every edge is adjacent to some other edge. So edge e shares an end with some other edge f in G, but the 2–point minors of G do not necessarily determine which end (unlike the case for directed graphs). If edge f is adjacent only to edge e then it makes no difference (as far as equality of edge–graphs is concerned) which end of f is incident with e. If edge f is adjacent to some other edge g say, then since g is not parallel to f, it is incident with exactly one end of f and so the 3–point minor $G \cdot \{e,f,g\}$ determines which end of f (the end incident with g, or the other one) is incident with e in G. Let one end of edge e be attached to this end of edge f in $G \backslash e$. If e is a loop then G has already been reconstructed. So suppose e is a link. If h∈Q–{e,f}, then the 3–point minor Q·{e,f,h} determines whether or not h is incident with the end of e that is not incident with f. If there is no such edge h, then this end of edge e can be a new vertex added for this purpose. Otherwise, this end of e can be attached to the appropriate end of edge h, determined as for edge f. Thus G is uniquely determined by its proper minors, contradicting the assumption that it is a ψ–structure. Therefore, the edge–graphs in figure 3 are the only ψ–structures of $\mathscr{G}$(EG,1).

It is easily seen that there are no natural excluded minors of order strictly less than 3. Counting reveals that, up to isomorphism, there are 30 order–3 structures in $\overline{\mathscr{G}(\text{EG–1})}$, the completion of $\mathscr{G}$(EG,1), 23 of which are in $\mathscr{G}$(EG,1) (that is, are 3–edge graphs) leaving 7 order–3 natural excluded minors of $\mathscr{G}$(EG,1). Now 7 is not too many to list, provided they can be described conveniently. (Remember, they are not in $\mathscr{G}$(EG,1), and are not graphs.) One way to describe a natural excluded minor is to list its *immediate minors*, that is, minors obtained by removal of a single point. (The number of these is the order of the structure multiplied by the number of manners of point removal.) Figure 4 illustrates the immediate minors of the seven natural excluded minors of order 3. It is assumed that the ground set (edge set) is {1,2,3} and in each row are the three immediate minors of a natural excluded minor, with the graph (with edge set {1,2,3}–{i}) obtained by deleting i, for each i∈{1,2,3}, in column i. Note that in any row, and for any distinct i,j∈{1,2,3}, deleting edge i from the graph in column j yields the same result as deleting edge j from the graph in column i, as required. Edges are labelled only when it is necessary to avoid ambiguity.

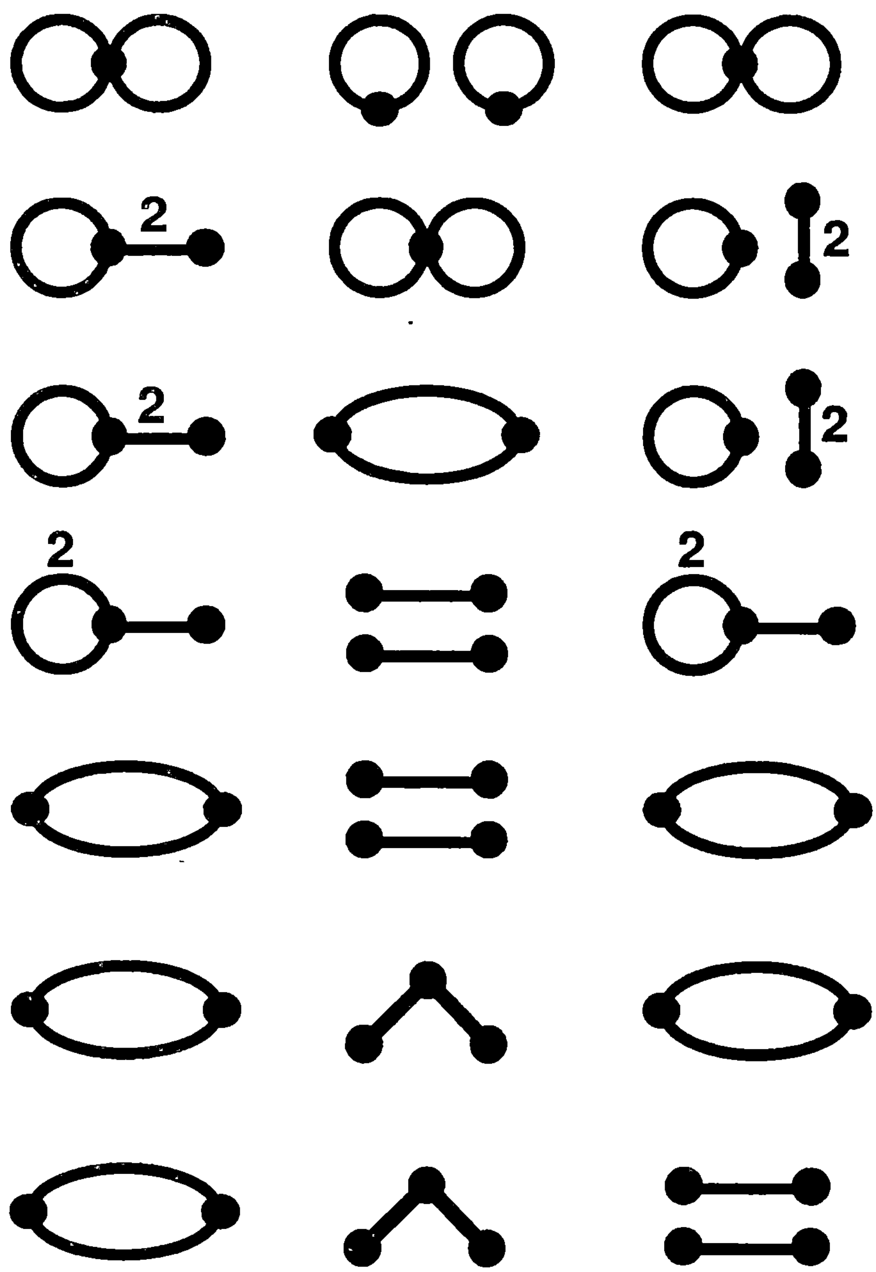

Figure 4: Immediate minors of 3–point
natural excluded minors of $\mathscr{G}$(EG,1).

Let us find the remaining natural excluded minors. Recall the proof that all the ψ–structures of $\mathscr{G}$(EG,1) have been found, but now suppose that G$\in\overline{\mathscr{G}(\text{EG,1})}$, with ground set Q is a natural excluded minor of $\mathscr{G}$(EG,1). Then G is not a graph, but if e$\in$Q, then G\e is a graph and the minors of G which are ψ–structures uniquely determine, for each edge in G\e, which ends of each such edge, e must be incident with. Since G is not a graph, this will be requiring that e does something inconsistent with G being a graph. If e is a loop (as determined by the minor G·e) then it either must be incident with two distinct vertices (involving exactly two other edges) or e must be both incident and not incident with some vertex (again involving two other edges, both incident with the vertex, but only one being adjacent to e). If e is parallel to some edge f say, (as determined by the minor G·{e,f}) then there must be exactly one other edge g which is connected to f differently to the way it is connected to e (so that G\e is not isomorphic to G\f via the bijection sending g to g and f to e). All these cases involve exactly three edges and are covered by figure 4. In the remaining case (where there are no loops or parallel edges) then either e must be incident with three vertices (involving exactly three other edges, not two by the exclusion of parallel edges) or e must be adjacent to some edge f say, but not adjacent to edges incident with either end of f (again involving three other edges). These cases all involve exactly four edges, and an exhaustive search (there are only finitely many order–4 structures in $\overline{\mathscr{G}(\text{EG,1})}$) yields 8 order–4 natural excluded minors. These are described by their immediate minors in figure 5. The comments referring to figure 4, also apply here except that the ground set is now {1,2,3,4}.

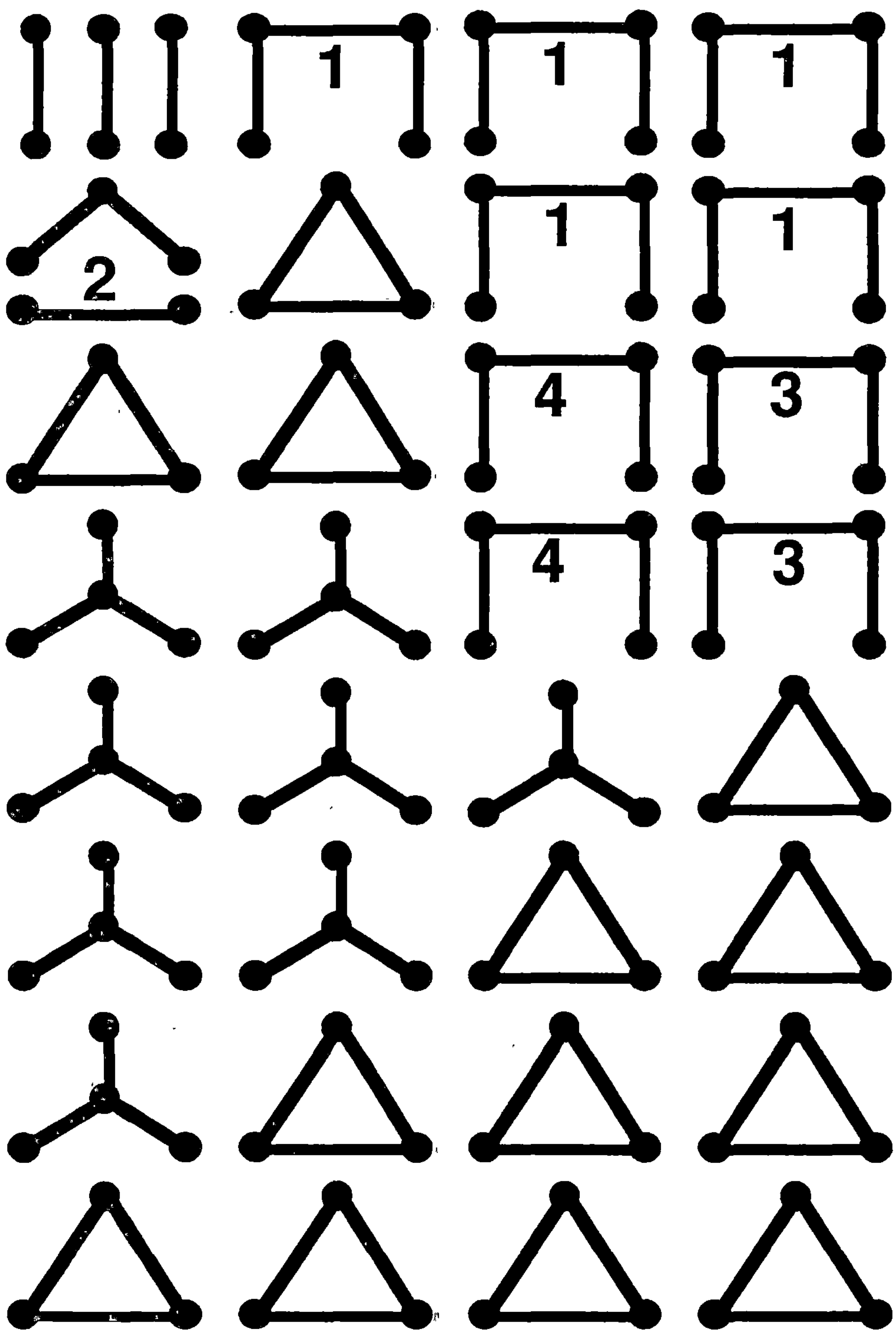

Figure 5: Immediate minors of 4–point

natural excluded minors of $\mathscr{G}$(EG,1).

The *edge colouring problem* is that of colouring the edges of a loopless edge–graph such that no two adjacent edges have the same colour. Construct a simple vertex–graph, with the same vertex set as the edge set of the edge–graph, and with two vertices adjacent (that is, they are the ends of a single link) exactly when the corresponding two edges are adjacent in the edge–graph. The edge colouring problem for the edge–graph is equivalent to the *vertex colouring problem* of the simple vertex graph, which is to colour the vertices such that no two adjacent vertices have the same colour. The function sending each loopless edge–graph in $\mathscr{G}$(EG,1)exc{loop}=$\mathscr{G}^{L}$(EG,1) say, to the corresponding simple vertex graph in $\mathscr{G}$(SVG) is in fact a homomorphism. This homomorphism is uniquely determined by specifying that it sends the two graphs on the left of row 4 in figure 3, to the 2–vertex simple vertex–graph with a single edge (link) and that it sends the graph on the right of row 4 in figure 3, to the 2–vertex simple vertex–graph without any edges. The only ψ–structures of $\mathscr{G}^{L}$(EG,1) are those in rows 4 and 5 of figure 2, and there remain three order–3 and eight order–4 natural excluded minors, as is immediate by observation. Let $\mathscr{G}^{E}$(SVG) be the homomorphic image of $\mathscr{G}^{L}$(EG,1) under this homomorphism. The excluded minors of $\mathscr{G}^{E}$(SVG) in $\mathscr{G}$(SVG) illustrated in figure 5. (They are all, of course, simple vertex graphs.) They are also its natural excluded minors, since $\mathscr{G}$(SVG) is complete, and $\mathscr{G}^{E}$(SVG) contains its core, so that $\mathscr{G}$(SVG) is the completion of $\mathscr{G}^{E}$(SVG)). The proof that these are the only ones involves a routine case analysis, but is too long (about as long as section 11) to include here.

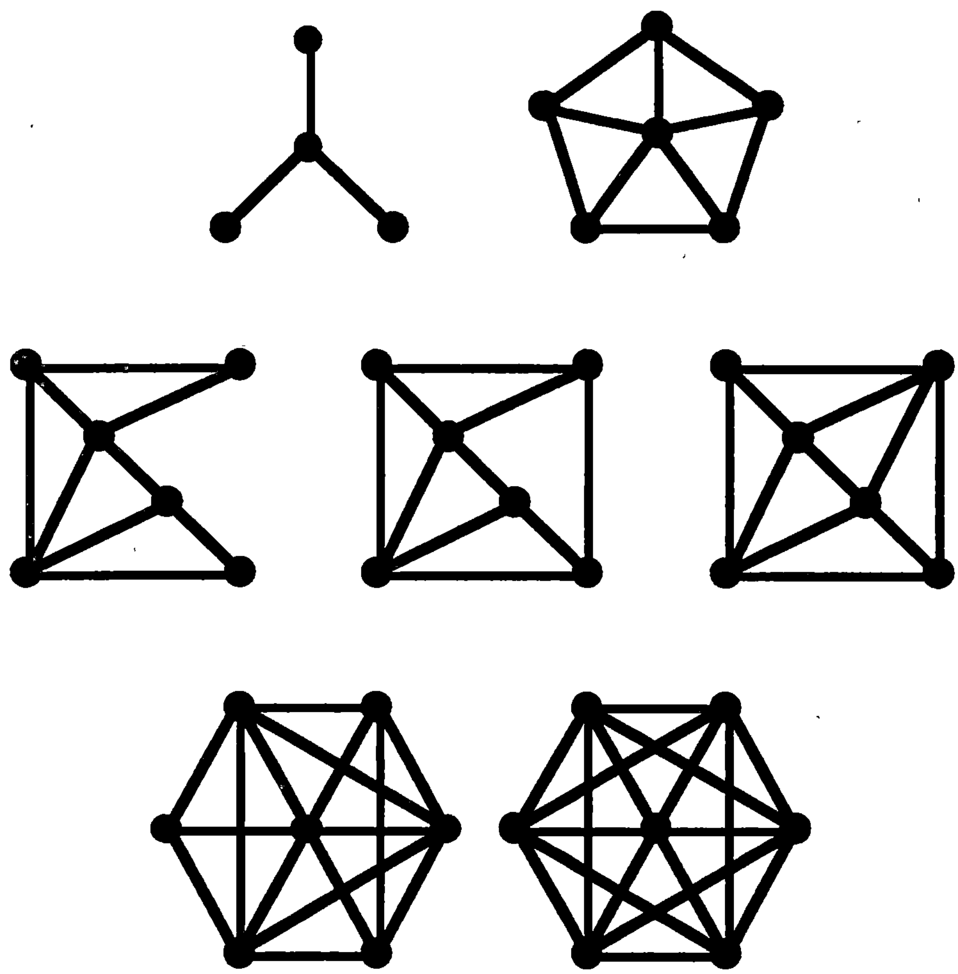

Figure 6: The natural excluded minors of $\mathscr{G}^{E}$(SVG).

The minor class $\mathscr{G}$(ED,2) has the same structures and ground sets as $\mathscr{G}$(ED,1), and it has edge deletion as before, but it also has a second manner of point removal called *edge contraction.* For any edge–digraph G∈$\mathscr{G}$(ED,2) and any edge e in its ground set E(G), the *contraction* of e from G is denoted G/e, and is obtained from G as follows. If e is a link then its two ends (head and tail) are identified into a single vertex (so that any edge with its head or tail being also an end of edge e, now has the single vertex as its head or tail) making edge e a loop (if it has not already one) which is then deleted. One way to visualise this is to consider the drawing of a graph. If an edge is contracted, the corresponding curve in the drawing of a graph is shrunk to a point so that its two ends merge into one, dragging any incident edges with them. For any $P \subseteq E(G)$, the graph G/P is obtained from G by contracting every edge in P (in any order).

While there is an obvious relationship between $\mathscr{G}$(ED,2) and $\mathscr{G}$(ED,1) it should be noted that, as unary algebras, they have different unary signatures, and so are in different varieties. Much, but not all, of universal algebra theory works within varieties, and cannot be used to connect these two minor classes. Nevertheless, they are connected by a signature modifying construction defined in section 6, namely that $\mathscr{G}$(ED,1) is $\mathscr{G}$(ED,2) confined to deletion, denoted $\mathscr{G}(\mathrm{ED},2) \mid \{\mathrm{deletion}\}$. Via this connection, useful information about the ψ–structures and natural excluded minors of $\mathscr{G}$(ED,2), is obtained from those of $\mathscr{G}$(ED,1).

If two edge–digraphs are ψ–equivalent in $\mathscr{G}$(ED,2), then all their corresponding proper minors are equal, in particular those obtained purely by deletion, so that they are ψ–equivalent in $\mathscr{G}$(ED,1). Therefore, to find the ψ–structures of $\mathscr{G}$(ED,2), it is only necessary to consider those of $\mathscr{G}$(ED,1), (see figure 2) testing them pairwise for ψ–equivalence. Neither of the two to the left of the dotted line in row 4 of figure 2 is ψ–equivalent to any of the five on the right. Taking the dotted line to split row 4 into two separate rows, figure 2 illustrates all the ψ–structures of $\mathscr{G}$(ED,2) with ψ–equivalent graphs in the same row.

Let us find an upper bound for the order of natural excluded minors of $\mathscr{G}$(ED,2). Let G be a natural excluded minor of $\mathscr{G}$(ED,2) and let its ground set Q have cardinality at least 3. Since $\overline{\mathcal{G}(\mathrm{ED},1)}$ (note it is 1, not 2) is complete, there is a structure $H\in\overline{\mathcal{G}(\mathrm{ED},1)}$, with ground set Q, such that $G\cdot\{e,f\}=H\cdot\{e,f\}$ for all distinct $e,f\in Q$. (Note that every $G\cdot\{e,f\}$ is indeed a graph.) There is only one possibility for H, since all the ψ-structures (and hence, all structures in the core) of $\mathcal{G}(\mathrm{ED},1)$ have order at most 2. For every proper subset P of Q, it follows that $G\cdot P=H\cdot P$, and these are graphs. Now G is not a graph, but H could be. If H is not a graph, then H is a natural excluded minor of $\mathcal{G}$ (ED,1) so that $|Q|=3$, and G is of order 3. If H is a graph, then H is a structure in $\mathcal{G}(\mathrm{ED},2)$ so that edges can be contracted from it. Clearly $G\neq H$, and since structures in $\overline{\mathcal{G}(\mathrm{ED},2)}$ are uniquely determined by their order–2 minors (the ψ-structures have order at most 2) there must exist some $P\subseteq Q$, and distinct edges e and f in Q–P such that $(G/P)\cdot\{e,f\}\neq(H/P)\cdot\{e,f\}$. Choose P so that $|P|$ is minimal with respect to this property, and choose $p\in P$. (Clearly $|P|\geq 1$.) Let $G'=(G/P-\{p\})\cdot\{p,e,f\}$ and let $H'=(H/P-\{p\})\cdot\{p,e,f\}$. So G' and H' are structures in $\overline{\mathcal{G}(\mathrm{ED},2)}$ with ground set $\{p,e,f\}$ and $H'\in\mathcal{G}(\mathrm{ED},2)$. For any distinct $q,r\in\{p,e,f\}$ it follows by the minimality of $|P|$ that $G'\cdot\{q,r\}=H'\cdot\{q,r\}$. If G' were a graph, then as above, G' would be H'. But $G'/p=(G/P)\cdot\{e,f\}\neq(H/P)\cdot\{e,f\}=H'/p$, so that $G'\neq H'$. Therefore G' is not a graph, and hence not in $\mathcal{G}(\mathrm{ED},2)$. By the minimality of G, it must be that $G=G'$ and $Q=\{p,e,f\}$, so that $|Q|=3$. Whether or not H is a graph, it follows that the order of G is 3. Therefore, every natural excluded minor of $\mathcal{G}(\mathrm{ED},2)$ has order at most 3. These are easily counted, and this is done below.

The two 1–edge digraphs (see row 1 of figure 2) can be named loop and coloop in the obvious way. It is then easily checked that $\mathcal{G}(\mathrm{ED},2)$ has the same six order–2 natural excluded minors as $\mathcal{M}$. The unique homomorphism from $\mathscr{G}$(ED,2) to $\mathscr{M}$, specified by sending loop to loop and coloop to coloop, is of great interest. It sends each edge–digraph to a *graphic matroid.* The homomorphic image of $\mathscr{G}$(ED,2) is $\mathscr{M}^{G}$, the minor class of graphic matroids. It is well known to matroid theorists that $\mathscr{M}^{G}$ has five excluded minors in $\mathscr{M}$, one each of orders 4,9 and 10 and two of order 7 [17]. Adding to this list, the (six order–2) natural excluded minors of $\mathscr{M}$, gives those of $\mathscr{M}^{G}$. Returning to the natural excluded minors of $\mathscr{G}$(ED,2); it is a simple matter to count those of order 3, but there are too many to list here, although enthusiastic readers may like to list them. There are 4890. This number is so much greater than for $\mathscr{G}$(ED,1) simply because having two manners of point removal creates many more possibilities. Some more detail can be provided by recalling, see section 12, that the above homomorphism can be extended to all of $\overline{\mathscr{G}(\mathrm{ED},2)}$, (with homomorphic image $\overline{\mathscr{M}}$). Of the order–3 natural excluded minors, this homomorphism sends 17 to U_0^3, 353 to $U_0^2+U_1^1$, 79 to $U_0^1+U_1^2$, 18 to U_1^3, 198 to U_2^3, 472 to $U_1^2+U_1^1$, 1040 to $U_0^1+U_2^2$ and 2713 to U_3^3. (The number of digraphs sent to these are 3,7,5,2,2,8,13 and 12 respectively.)

The minor class $\mathscr{G}$(EG,2) has the same structures and ground sets as $\mathscr{G}$(EG,1), and it has edge deletion as before, but it also has a second manner of point removal, namely edge contraction, defined and denoted as it is for $\mathscr{G}$(ED,2). So there is a surjective homomorphism from $\mathscr{G}$(ED,2) to $\mathscr{G}$(EG,2) which, as usual, sends each edge–digraph to the corresponding edge–graph by removing the direction of each edge. The homomorphism can be metaphorically extended to the whole discussion about $\mathscr{G}$(ED,2), yielding the following abbreviated discussion.

Again $\mathscr{G}$(EG,1) is $\mathscr{G}$(EG,2) | {deletion}. The ψ–structures of $\mathscr{G}$(EG,2) are among those of $\mathscr{G}$(EG,1), see figure 3. Figure 7 illustrates these with the usual conventions.

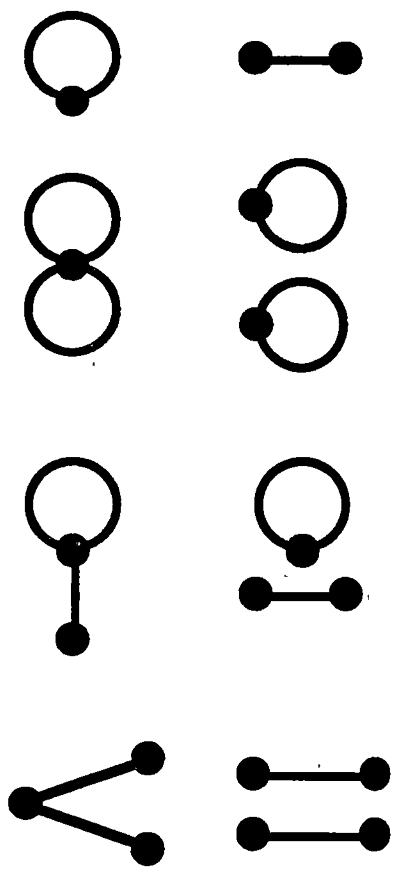

Figure 7: The ψ–structures of $\mathscr{G}$(EG,2).

An argument given earlier deduces that the natural excluded minors of $\mathscr{G}$(ED,2) have order at most 3, from the fact that those of $\mathscr{G}$(ED,1) have order at most 3, while the ψ–structures of $\mathscr{G}$(ED,1) have order at most 2. This can be adapted, or generalised, to deduce that the natural excluded minors of $\mathscr{G}$(EG,2) have order at most 4, from the fact that those of $\mathscr{G}$(EG,1) have order at most 4, while the ψ–structures of $\mathscr{G}$(EG,1) have order at most 3. Again naming the two 1–edge graphs (see figure 7) loop and coloop, in the obvious way, it is easily checked that $\mathscr{G}$(EG,2) has the same six order–2 natural excluded minors as $\mathscr{M}$. There is a surjective homomorphism from $\mathscr{G}$(EG,2) to $\mathscr{M}^G$, just as there was from $\mathscr{G}$(ED,2). (The one from $\mathscr{G}$(ED,2) to $\mathscr{M}^G$ is the composition of that from $\mathscr{G}$(ED,2) to $\mathscr{G}$(EG,2) with that from $\mathscr{G}$(EG,2) to $\mathscr{M}^G$.) Minor class $\mathscr{G}$(EG,2) has 125 3–point (these are easily counted) and 6 4–point (these have to be "found") natural excluded minors. The above homomorphism, extended (uniquely) to $\overline{\mathscr{G}(\mathrm{EG},2)}$, sends 17 of these to U_0^3, 35 to $U_0^2+U_1^1$, 8 to $U_0^1+U_1^2$, 3 to U_1^3, 3 to U_2^3, 8 to $U_1^2+U_1^1$, 35 to $U_0^1+U_2^2$, 16 to U_3^3, 1 to U_2^4, 2 to U_3^4 and 3 to U_4^4. (The number of graphs sent to those is 3,5,2,1,1,2,5,4,0,1 and 8 respectively.)

The next example is of interest because it involves the direct product of two minor classes in an uncontrived way. It relates to the drawing of graphs on surfaces. There are several distinct approaches to this problem. The approach here resembles that (for the plane) in [18], but is quite different to that in [16]. An *embedding* of a graph G on a (locally Euclidean) surface Γ, is a drawing of G on Γ for which none of the curves representing edges intersect. If such an embedding exists, then G is *embeddable* on Γ. Two graphs G and H are *compatible* on Γ if they have the same edge set, are both embeddable on Γ, and they can be drawn on Γ where the only intersections are as follows; for each edge e, the curve representing e in G, crosses over the curve representing e in H (and visa versa) at exactly one point. Given a graph G embeddable in Γ, a graph H can be constructed, such that G and H are compatible on Γ, as follows. A drawing of G on Γ partitions the points on Γ, which are not on a vertex or edge, into faces. Let H have the same edge set as G, and have the faces as its vertices, where each edge is incident with the two (not necessarily distinct) faces on either side of it on the surface. Whenever two graphs G and H are compatible on Γ, it is easily shown that $G\backslash e$ and H/e are compatible on Γ, and G/e and $H\backslash e$ are compatible on Γ. (Also $G\backslash e$ and $H\backslash e$ are compatible on Γ, but that is not used here.) This inspires the minor class $\mathscr{G}$(PEG,2) defined below.

Let the minor class $\mathcal{G}^*$(EG,2) be obtained from $\mathcal{G}$(EG,2) by swapping deletion and contraction. (See section 6.) That is, edge deletion in $\mathcal{G}^*$(EG,2) is the same as edge contraction in $\mathcal{G}$(EG,2), and visa versa. The ψ–structures of $\mathcal{G}^*$(EG,2) and $\mathcal{G}$(EG,2) are the same, and there is an obvious one–to–one correspondence between their natural excluded minors. Let the minor class of *pairs of edge–graphs,* G(PEG,2), be the direct product $\mathcal{G}$(EG,2)×$\mathcal{G}^*$(EG,2). The structures of $\mathcal{G}$(PEG,2) are pairs of edge–graphs (G,H), where G and H have the same edge set. For each edge e in this edge set, deleting e from (G,H) yields (G\e,H/e), and contracting e from (G,H) yields (G/e,H\e). This minor class has one mixed automorphism, other than the identity, namely that which swaps deletion and contraction, and which sends each pair (G,H) to (H,G), the *dual* of (G,H). (Note that the only mixed automorphism of $\mathcal{G}$(EG,2) is the identity.) The ψ–description of $\mathcal{G}$(PEG,2) is routinely determined from the ψ–descriptions of $\mathcal{G}$(EG,2) and $\mathcal{G}^*$(EG,2), as shown in section 11. On any 1–element ground set there are four structures which are ψ–equivalent (since there is a unique order–0 structure.) On any 2–element ground set, $\mathcal{G}$(EG,2) (and $\mathcal{G}^*$(EG,2)) has nine structures which ψ–equivalence partitions into 4 lots of 2 and 1 lot of 1 (see figure 7, and consider row 3 turned upside–down) so that $\mathcal{G}$(PEG,2) has 81 structures which ψ–equivalence partitions into 16 lots of 4, 8 lots of 2, and 1 lot of 1. (Up to isomorphism there are 52 order–2 ψ–structures, in 10 lots of 4 and 6 lots of 2. This statement makes sense, since there are no distinct, but isomorphic, ψ–equivalent structures.) Also on any 2–element ground set $\mathcal{G}$(EG,2) (and $\mathcal{G}^*$(EG,2)) has 11 natural excluded isominors (see figure 1), only one of which has the bijection swapping the two ground set elements as an automorphism, so that there are (11+1)/2=6 natural excluded minors, by Polya enumeration. It follows that $\mathcal{G}$(PEG,2) has $(11^2+1^2)/2=61$ natural excluded minors of order 2. Similarly, since $\mathscr{G}$(EG,2) (and $\mathscr{G}^*$(EG,2)) has (558+60+60+60+6+6)/6=125 natural excluded minors of order 3, it follows that $\mathscr{G}$(PEG,2) has $(558^2+60^2+60^2+60^2+6^2+6^2)/6=53706$ natural excluded minors of order 3.

For any (locally Euclidean) surface Γ, let $\mathscr{G}^\Gamma$(PEG,2) be the sub minor class (as earlier arguments show it is) of $\mathscr{G}$(PEG,2), consisting of exactly those pairs (G,H) such that G and H are compatible on Γ. Consider the homomorphism from $\mathscr{G}^\Gamma$(PEG,2) into $\mathscr{G}$(EG,2) which sends each pair (G,H) to G, and let $\mathscr{G}^\Gamma$(EG,2) be the homomorphic image. Earlier argument shows that $\mathscr{G}^\Gamma$(EG,2) is the sub minor class of $\mathscr{G}$(EG,2) consisting of exactly those edge–graphs which are embeddable on Γ. More knowledge about the excluded minors of $\mathscr{G}^\Gamma$(PEG,2) in $\mathscr{G}$(PEG,2), and about minor class homomorphisms, could yield information about the excluded minors of $\mathscr{G}^\Gamma$(EG,2) in $\mathscr{G}$(EG,2). Clearly $\mathscr{G}^\Gamma$(PEG,2) has the mixed automorphism described earlier for $\mathscr{G}$(PEG,2), so that if (G,H) is an excluded minor, then so is (H,G). I conjecture that there is a bound on the order of these excluded minors, approximately proportional to the genus [16] of the surface Γ.

When Γ is the plane, then an order–1 excluded minor is the pair of graphs, where the single edge in each graph is a loop. An excluded minor of order 3 is given by the pair of graphs depicted in figure 8.

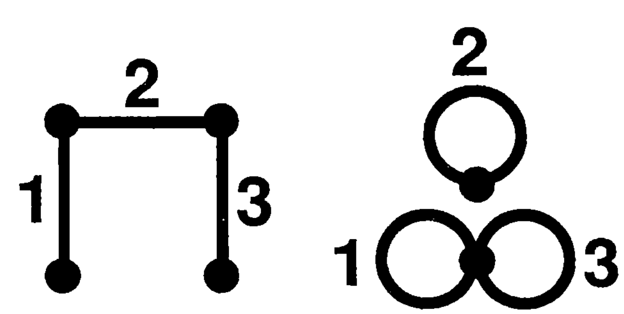

Figure 8: An excluded minor of $\mathscr{G}^\Gamma$(PEG,2) in $\mathscr{G}$(PEG,2) where Γ is the plane.

It is easily seen that this pair is not compatible on the plane, despite the fact that they are both embeddable on the plane. (Consider the left hand graph and edges 1 and 3 of the right hand graph drawn subject to compatibility. Edge 2 of the latter graph cannot be added subject to compatibility.) By observation, there are six pairs of this nature, giving rise to twelve order–3 excluded minors (since the graphs can appear in either order). I conjecture that there are no other excluded minors.

SECTION 15: IDEAS FOR FURTHER CONSIDERATION

This final section presents some ideas which could form a basis for further research. Standards of rigour are dropped for this discussion. (Many of the claims made are not difficult to prove, but while they are of interest, the proofs would not justify the space they would occupy at this stage. There are also conjectures which there is plenty of motivation to attempt to prove.) Again our attention is restricted to minor classes with finite ground sets, as in section 11. (Note that minor classes with infinite ground sets enable us to talk about minor classes of matroids or graphs with such ground sets, which may be worth considering.)

It should be apparent that a potentially rewarding area of study is the study of minor class homomorphisms. A desirable result would be one of the following form. If α is a minor class homomorphism from $\mathscr{S}$, and α and $\mathscr{S}$ satisfy certain conditions, then $\alpha(\mathscr{S})$ has finitely many natural excluded minors, or better still, there is a certain upper bound on the order of these natural excluded minors. Many seemingly intractable combinatorial problems would be solved by such a theorem, (although the proof of such a theorem may itself seem to be an intractable combinatorial and algebraic problem). Clearly the conditions on α and $\mathscr{S}$ should include that certain quantities associated with α and $\mathscr{S}$ are finite, (since otherwise one could probably construct a counterexample in which $\alpha(\mathscr{S})$ had infinitely many natural excluded minors). The important examples in sections 12 to 14 all satisfy these finiteness conditions. Firstly $|K|$, the number of manners of point removal, should be finite. (It is usually 1 or 2.) Secondly $\mathscr{S}$ should have finitely many ψ-structures up to isomorphism. Equivalently, when $|K|$ is finite, the core $\mathscr{S}^{\psi}$ is *finitely generated* (that is, generated by finitely many structures), or each ψ-equivalence class is finite and the order of ψ-structures is bounded, and these imply that $|\mathscr{S}_Q|$ is finite for every ground set Q. Let $\mathscr{T}=\alpha(\mathscr{S})$.

Thirdly the core $\mathcal{S}^{\psi}$ should also be finitely generated. (Recall that α is expressed as the unique homomorphism which extends $\alpha|^{\mathcal{S}^{\psi}}$, and $\alpha|^{\mathcal{S}^{\psi}}$ need only be specified on finitely many values in $\alpha^{-1}(\mathcal{S}^{\psi})$, since the above conditions guarantee that $\alpha^{-1}(\mathcal{S}^{\psi})$ is finitely generated.) Fourthly $\mathcal{S}$ should have finitely many natural excluded minors. When the second and fourth conditions hold, *$\mathcal{S}$ has a finite ψ–description.* However these are not sufficient to ensure that $\alpha(\mathcal{S})$ has finitely many natural excluded minors, as the following example shows.

Let $\mathcal{W}^{\downarrow}$ be the sub minor class of $\mathcal{W}$ whose structures on ground set Q are $(\emptyset,Q)$ and $(\{\emptyset\},Q)$. Let $\mathcal{W}^{\uparrow}$ be the sub minor class of $\mathcal{W}$ whose structures on ground set Q are $(\emptyset,Q)$ and $(\{Q\},Q)$. The core of these is $\mathcal{W}^{\psi}$, since the only ψ–structures are $(\emptyset,\emptyset)$ and $(\{\emptyset\},\emptyset)$. Observing that $\mathcal{W}$ is complete, the natural excluded minors of $\mathcal{W}^{\downarrow}$ are $(\{\{q\}\},\{q\})$ and $(\{\emptyset,\{q\}\},\{q\})$ while those of $\mathcal{W}^{\uparrow}$ are $(\{\emptyset\},\{q\})$ and $(\{\emptyset,\{q\}\},\{q\})$ for some 1–element ground set $\{q\}$. Therefore $\mathcal{W}^{\downarrow}$ and $\mathcal{W}^{\uparrow}$ both have a finite ψ–description, and it is routine to show that their disjoint union also has a finite ψ–description. However a homomorphic image of this disjoint union is the union $\mathcal{W}^{\downarrow} \cup \mathcal{W}^{\uparrow}$ which has a finitely generated core $\mathcal{W}^{\psi}$, but has infinitely many natural excluded minors. Choosing a ground set Q of cardinality n, for each $n \in \{2,3,\cdots\}$, every natural excluded minor is of the form $(\{\emptyset,\{Q\}\},Q)$.

Another condition is needed which excludes the above counterexample, but not the examples in sections 12 to 14. An *(ascending) chain* in a minor class $\mathcal{S}$ is a sequence of structures $S_1,S_2,S_3,\cdots$ (usually, but not necessarily, infinite) such that $S_1 \leq S_2 \leq S_3 \leq \cdots$ in the quasi order of $\mathcal{S}$. The minor class $\mathcal{S}$ *is generated by a chain* if $\mathcal{S} = \mathcal{S}\,\mathrm{inc}\{S_1,S_2,S_3,\cdots\}$ for some chain $S_1,S_2,S_3,\cdots$. (Note that in this case, $\alpha(\mathcal{S})$ is generated by the chain $\alpha(S_1),\alpha(S_2),\alpha(S_3),\cdots$.) Equivalently when $|K|$ is finite and $\mathcal{S}$ has a finite ψ–description, if $\mathcal{S}$ is the union of two sub minor classes, then one of them is $\mathscr{S}$, or for any two structures in $\mathscr{S}$, there is a structure having both of them as isominors, or for any sub minor class of $\mathscr{S}$ whose structures have bounded order, there is a structure in $\mathscr{S}$ which has each structure in the sub minor class as an isominor. It is easily arranged that each structure S_i in the chain is of order i, and that S_i is obtained from S_{i+1} by removing a single point in some manner. If some manner is used only finitely many times in this way, then it is easily arranged that it is used zero times. For any non-empty subset $K' \subseteq K$ define a K′-chain to be a chain of the latter form where each manner of point removal in K′ is used infinitely many times. (There are $2^{|K|}-1$ possibilities.) One can define a minor class freely generated by a K′-chain so that its homomorphic images are exactly the minor classes generated by a K′-chain. The important minor classes in section 12 and 14 are all generated by a chain, and those with K={delete, contract} are generated by a deletion-chain, as is easily verified. To the previous four conditions on α and $\mathscr{S}$, add the fifth condition that $\mathscr{S}$ is generated by a chain. I know of no counterexample to the statement that if $|K|$ is finite, $\mathscr{S}$ has a finite ψ-description and is generated by a chain, and the core of $\alpha(\mathscr{S})$ is finitely generated, then $\alpha(\mathscr{S})$ has finitely many natural excluded minors (and hence $\alpha(\mathscr{S})$ has a finite ψ-description). I conjecture that this statement is true, which is not to say that I necessarily believe the statement. Many important examples are covered by the case when K={delete,contract} and $\mathscr{S}$ is generated by a deletion-chain, and it is possible that the conjecture holds here but not generally. It may be that more conditions are needed, or worse still, that there is no conveniently stated theorem of the above form so that each example must be dealt with individually (and the ψ-description of $\alpha(\mathscr{S})$ found "by hook or by crook" in each case).

Let us make some observations about the minor class homomorphisms in sections 12 to 14. The homomorphic image usually has less ψ–structures (up to isomorphism) than the original minor class, and although this is not necessary, it can always be arranged, as shown later. Generally the minor class has many natural excluded minors with low order, which are easily found, while the homomorphic image has fewer natural excluded minors of higher order, which are more difficult to find. For example $\mathscr{G}$(ED,2) (whose structures are edge–digraphs) has 6 2–point and 4890 3–point natural excluded minors, and these are easy to count. In fact it is easy to prove that this list is complete, and such a proof effectively describes the structures in the list. The homomorphic image $\mathscr{G}$(EG,2) (whose structures are edge–graphs) has 6 2–point, 125 3–point and 6 4–point natural excluded minors. Finding those of order 4, and showing that these are the only ones, is clearly the major part of the task. The homomorphic image $\mathscr{M}^{\mathrm{G}}$ (whose structures are graphic matroids) of $\mathscr{G}$(EG,2) (and $\mathscr{G}$(ED,2)), has 6 2–point, 1 4–point, 2 7–point, 1 9–point and 1 10–point natural excluded minor. Finding these and showing they are the only ones is much harder than the corresponding problem for $\mathscr{G}$(EG,2) and $\mathscr{G}$(ED,2). These three minor classes all have two order–1 ψ–structures while $\mathscr{G}$(ED,2) has twelve, $\mathscr{G}$(EG,2) has six, and $\mathscr{M}^{\mathrm{G}}$ has zero order–2 ψ–structures. It is not surprising that the ψ–description of a minor class becomes substantially harder to find and verify when structures of high order are involved (despite there being far fewer structures involved) since the difficulty of dealing with structures increases rapidly with their order. However, the above behaviour would be highly desirable, were it to occur in general. In fact our handful of examples make it seem feasible that from α and $\mathscr{S}$, an upper bound could be found for the order of natural excluded minors of $\alpha(\mathscr{S})$, from which the ψ–description of $\alpha(\mathscr{S})$ could be routinely found. Of course, a handful of examples is no evidence for anything. It is more likely that our examples have other desirable properties, yet to be discovered.

In many of the examples, $\mathscr{S}$ and α are such that the only sub minor class of $\mathscr{S}$ whose homomorphic image under α is $\alpha(\mathscr{S})$, is $\mathscr{S}$ itself. This is the case for the homomorphism from $\mathscr{G}$(EG,2) to $\mathscr{M}^{G}$ by the following reasoning. The (undirected) graph K_n has n vertices, and exactly one edge between each pair of vertices (so that there are $\binom{n}{2}$ edges). Excluding any graph from $\mathscr{G}$(EG,2) excludes K_n, for some n, and K_n is the unique graph sent to the corresponding matroid, so that this matroid is excluded from $\mathscr{M}^{G}$. Actually this homomorphism satisfies the stronger property that $\mathscr{G}$(EG,2) is generated by the chain $K_1, K_2, K_3, \cdots$ and each K_n is the unique graph sent to the corresponding matroid. The homomorphisms from $\mathscr{G}$(ED,2) to either $\mathscr{G}$(EG,2) or $\mathscr{M}^{G}$ satisfy the weaker property but not the stronger.

The homomorphism from $\mathscr{G}$(ED,2) to $\mathscr{G}$(EG,2) has another interersting property. Recall from section 5 that a congruence on a unary algebra can be constructed from any group of automorphisms of that algebra. Recall from section 11 that an automorphism of a minor class $\mathscr{S}$, restricts to an automorphism of its core $\mathscr{S}^{\psi}$, and when it is extended to an automorphism of its completion $\bar{\mathscr{S}}$, the latter automorphism permutes the natural excluded minors of $\mathscr{S}$, (and similarly when $\mathscr{S}$ is treated as a (proper) pseudo minor class). Treating $\mathscr{G}$ (ED,2) as a (proper) pseudo minor class, each automorphism corresponds to a set P of ground set elements (not necessarily finite, even though each ground set is) such that each edge–digraph is sent to the edge–digraph with the direction of the edges in P reversed. (This assumes that the ground sets are exactly the finite subsets of some infinite set, an assumption made in section 11.) The corresponding congruence clearly respects isomorphism (so that it is a minor class congruence of $\mathscr{G}$(ED,2)) and is in fact the kernel of the homomorphism from $\mathscr{G}$(ED,2) to $\mathscr{G}$(EG,2).

It is interesting to consider this construction applied to $\mathscr{D}$(R) when R is a field. First consider the automorphisms of $\mathscr{D}$(R) considered as a minor class (which, when extended to $\overline{\mathscr{D}(R)}$, permute the natural excluded minors, as well as the ψ-structures). By consideration of the natural excluded minors, each automorphism sends loop to loop, coloop to coloop, and for some field automorphism β of R, sends each slope (a,q,r) (see section 13) to (β(a),q,r), as is routinely shown. (The fact that β respects field multiplication follows from the 3-point natural excluded minors, while respecting addition follows from those of order 4.) Treating $\mathscr{D}$(R) as a (proper) pseudo minor class, each automorphism is the composition of one of the above automorphisms, and one which for some function w, from all ground set elements to R, sends loop and coloop to themselves, and sends each slope (a,q,r) to (w(q)a/w(r),q,r). (The composite automorphism sends each subspace S of R^Q to the subspace $\{(w(q)\beta(x_q)\,|\,q\in Q)\,|\,(x_q\,|\,q\in Q)\in S\}$ of R^Q.) The corresponding congruence is intimately related to the topic of *equivalent representations.* Two *representations* f:Q→S and g:Q→T, where S and T are vector spaces over R and Q is a ground set, are equivalent if one can be obtained from the other by a combination of the following four elementary equivalences, (1) if S is a subspace of T and f(q)=g(q) for every q∈Q, (2) if there is a vector space isomorphism ω:S→T such that g=ω∘f, (3) if S=T and there is a field automorphism β of R such that $\overline{\beta}$:S→S is the identity on some basis of S and extends to S in the obvious way induced by β , then g=$\overline{\beta}$∘f, (4) if there is a function w:Q→R such that g(q)=w(q)f(q) for every q∈Q. The first two are accounted for by considering closure operators (matroids) σ_f and σ_g, and the latter two are accounted for by the above congruence. It is well known that this congruence is the kernel of the homomorphism from $\mathscr{D}$(R) to $\mathscr{M}$(R), only when |R|=2 or 3, otherwise the congruence is strictly below the kernel in the lattice of congruences of $\mathscr{D}$(R). It can be deduced from [8] that when |R|=4, the homomorphic image of $\mathscr{D}$(R) associated with this congruence, has ψ-structures of order 1 and 8, and another non-trivial homomorphism is required to obtain

$\mathscr{M}$(R). This seems to worsen the problem of finding the ψ-description of a homomorphic image to that of finding the ψ-description of a homomorphic image of a homomorphic image. Nevertheless, equivalent representations are widely studied and so there must be some merit in further considering the above material.

A desirable property which may be possessed by a minor class is that it is *well quasi ordered,* that is, the corresponding quasi order is a well quasi order, which means that there are no infinite descending chains (see section 5) and no infinite antichains (sequences $S_1,S_2,S_3,\cdots$ such that there are no distinct i and j such that $S_i \leq S_j$). Several equivalent conditions on the quasi order or the corresponding lattice of sub minor classes are given in [7]. For example, the minor class $\mathscr{S}$ is well quasi ordered if and only if every sub minor class of $\mathscr{S}$ has finitely many excluded minors in $\mathscr{S}$. Note that the property of being well quasi ordered is independent of the property of having finitely many natural excluded minors. For example, the empty minor class has both properties, $\mathscr{W}^{\downarrow} \cup \mathscr{W}^{\uparrow}$ has only the first, $\mathscr{W}$ has only the second, and $\mathscr{W}\,\mathrm{exc}(\{(\{\{q\}\,|\,q \in Q\} \cup \{\{Q\}\},Q)\,|\,Q \in \mathscr{Q} \text{ and } |Q|>2\})$ has neither property. Showing that a minor class is well quasi ordered is generally very difficult. In a lengthy series of papers previewed in [11], Robertson and Seymour have shown that $\mathscr{G}$(EG,2) is well quasi ordered. Clearly if a minor class is well quasi ordered, then so are all its homomorphic images and sub minor classes, so in particular $\mathscr{M}^{G}$ is well quasi ordered. In private correspondence, Robertson has conjectured that $\mathscr{M}$(R) is well quasi ordered whenever R is a finite field. One might conjecture more generally that $\mathscr{G}$(ED,2) is well quasi ordered, as well as $\mathscr{D}$(R) for every finite field R. To prove such results, one would need results to construct "larger" well quasi ordered minor classes from "smaller" ones, as Robertson and Seymour seem to have done. However, any such results would not be restricted to minor classes, and in fact minor classes probably contribute little to the study of well quasi ordering apart from unifying that part of the latter topic which intersects the former. Unfortunately, the direct product of two well quasi ordered minor classes need not be well quasi ordered, since $\mathscr{W}^{\downarrow}\times\mathscr{W}^{\uparrow}$ is not. (Observe that this minor class is also not generated by a chain, even though $\mathscr{W}^{\downarrow}$ and $\mathscr{W}^{\uparrow}$ are.) I know of no counterexample to the statement that if $|K|$ is finite and $\mathscr{S}$ has a finitely generated core, is generated by a chain, and is well quasi ordered, then $\mathscr{S}$ has finitely many natural excluded minors, so I conjecture it to be true.

Observe that in the variety of $(\mathscr{Q},K)$ minor class, the free minor classes, freely generated by one structure, are all well quasi ordered if and only if $\mathscr{Q}$ is a hereditary set of finite sets and K is finite. In this case, these free minor classes also have no infinite ascending chain. The conjecture about the ψ-description of homomorphic images could be extended to algebras in a special unary variety where all the free algebras, freely generated by one element, are well quasi ordered and have no infinite ascending chain.

Let us examine minor class homomorphisms more closely. There is no loss of generality in considering a congruence $\mathfrak{q}$ on $\mathscr{S}$, and the minor class $\mathscr{S}/\mathfrak{q}$. For each non-negative integer n, let the congruence $\mathfrak{q}_n$ be such that $S\mathfrak{q}_nT$ if and only if $S=T$ or $S\mathfrak{q}T$ and $|G(S)|=|G(T)|\leq n$. Clearly $\mathfrak{q}_0\leq\mathfrak{q}_1\leq\mathfrak{q}_2\leq\cdots$ and $\mathfrak{q}$ is the join of these. The sequence $\mathscr{S},\mathscr{S}/\mathfrak{q}_0,\mathscr{S}/\mathfrak{q}_1,\mathscr{S}\mathfrak{q}_2,\cdots$ converges to $\mathscr{S}/\mathfrak{q}$ in the following sense. First, for any minor class $\mathscr{T}$, let $\mathscr{T}|^n$ be the sub minor class consisting of all structures in $\mathscr{T}$ with order at most n. Then $(\mathscr{S}/\mathfrak{q})|^n$ is identical to $(\mathscr{S}/\mathfrak{q}_m)|^n$ whenever $m\geq n$ (in particular, when m=n). It follows that the ψ-structures of order at most n and the natural excluded minors of order at most n+1 of $\mathscr{S}/\mathfrak{q}_n$ are the same as those of $\mathscr{S}/\mathfrak{q}$, but those of higher order can be quite different.

Compare the ψ-descriptions of $\mathscr{S}/\mathfrak{q}_{n-1}$ and $\mathscr{S}/\mathfrak{q}_n$. There is a homomorphism $\beta_n:\mathscr{S}/q_{n-1}\to\mathscr{S}/q_n$ where $\beta_n(Sq_{n-1})=Sq_n$ for every $S\in\mathscr{S}$. (This is well defined since $q_{n-1}\leq q_n$.) Now β_n is the identity on structures whose order is not n, but may be non-injective on n-point structures. Suppose S and T are n-point structures in $\mathscr{S}/q_{n-1}$ with $\beta_n(S)=\beta_n(T)$ (so that, $G(S)=G(T)$. Then for any prescription $\mathfrak{K}\in\overline{K}^Q$ with $G(\mathfrak{K})\neq Q$ it follows that $S[\mathfrak{K}]=\beta_n(S[\mathfrak{K}])=(\beta_n(S))[\mathfrak{K}]=(\beta_n(T))[\mathfrak{K}]=\beta_n(T[\mathfrak{K}])=T[\mathfrak{K}]$, so that $S\psi T$ (although $S\psi T$ need not imply $\beta_n(S)=\beta_n(T)$). That is, distinct structures "clumped together" by β_n must be ψ-equivalent ψ-structures of order n. Now $\mathscr{S}/q_n$ has the same ψ-structures of order less than n, as $\mathscr{S}/q_{n-1}$, and those order n are determined by the fact that for structures S and T of order n in $\mathscr{S}/q_{n-1}$, $\beta_n(S)\psi\beta_n(T)$ exactly when $S\psi T$ (since β_n is the identity on $(\mathscr{S}/q_{n-1})|^{n-1}$). For structures of order greater than n, ψ-equivalence in $\mathscr{S}/q_n$ is the same as in $\mathscr{S}/q_{n-1}$ except that some new ψ-equivalence may be introduced among structures of order n+1. (Since two order n+1 structures S and T in both $\mathscr{S}/q_{n-1}$ and $\mathscr{S}/q_n$ may have immediate minors which are different in $\mathscr{S}/q_{n-1}$ but not in $\mathscr{S}/q_n$). This new ψ-equivalence may extend to (n+1)-point natural excluded minors of $\mathscr{S}/q_{n-1}$, and those which become ψ-equivalent to an (n+1)-point structure in $\mathscr{S}/q$ vanish, while natural excluded minors becoming ψ-equivalent, become equal. The only other change to the natural excluded minors is that new ones of order n+2 in $\mathscr{S}/q_n$ may appear. (Since "clumping together" structures of order n creates new possibilities for homomorphisms from $E(1^Q)$ to $\mathscr{S}/q_n$ when $|Q|=n+2$.)

Now consider how the ψ-description changes as we move along the sequence. Typically we know the ψ-structures of $\mathscr{S}/q$ (it is the natural excluded minors that need to be found) so that we know which structures each β_n "clumps together". When β_n is applied to obtain $\mathscr{S}/q_n$ from $\mathscr{S}/q_{n-1}$ some (n+1)-point natural excluded minors are "clumped together" while others vanish (those remaining, remain in $\mathscr{S}/q$) and some (n+2)-point natural excluded minors are created. Typically, most of these (n+2)–point natural excluded minors promptly vanish again when β_{n+1} is applied to give $\mathscr{S}/q_{n+1}$. "Most" becomes "all" if $\mathscr{S}/q$ has no natural excluded minors of order n+2, (in particular when all the natural excluded minors of $\mathscr{S}/q$ have order less than n+2.) The difficulty of characterising when this happens provides insight into why the ψ–description of $\mathscr{S}/q$ is so difficult to find. Even when $\mathscr{S}$ and q satisfy the desirable properties mentioned in the conjecture about homomorphic images, there seems to be no good reason why there should be finitely many natural excluded minors. Other approaches to this problem yield a similar situation, reinforcing the difficulty of the problem.

In attempting to prove the conjecture about homomorphic images, the above argument, and the argument in section 10 following theorem 10.4, shows that we can confine our attention to homomorphisms of the form $\Omega^{\mathscr{T},\mathscr{S}}:\mathscr{S}\to\bar{\mathscr{T}}$ where $\mathscr{T}$ is the core of $\bar{\mathscr{T}}$. (See section 10.) Such homomorphisms are the identity on $\mathscr{T}$, and the homomorphic image retains only those ψ–structures of $\mathscr{S}$, which are in $\mathscr{T}$. (In the patterned hypercube visualisation of section 11, this simply amounts to "blanking" the patterned faces corresponding to eliminated ψ–structures. The homomorphism from $\mathscr{D}$(R) to $\mathscr{M}$(R) is already of this form when R is a field.)

Readers wishing to examine this topic further should not be discouraged by the difficulty of the abovementioned conjecture. (It is only stated in a form which avoids known counterexamples, and is likely to need further refinement.) Rather than attempting such conjectures, it is more rewarding to examine the situation from different angles, providing insight into various aspects of the problem. The by–products of such investigations are usually of more consequence than the final result itself.

REFERENCES

1. G. Birkhoff, On the structure of abstract algebras, *Proc. Cambridge Philos. Soc.* **31** (1935), 433–454.
2. R.E. Bixby, On Reid's characterisation of ternary matroids, *J. Combinatorial Theory* (B) **26** (1979), 174–204.
3. H. Ehrig and B. Mahr, "Fundamentals of Algebraic Specification", Springer–Verlag, Berlin, 1985.
4. S. Eilenberg and S. MacLane, General theory of natural equivalences, *Trans. A.M.S.* **58** (1945), 231–294.
5. G. Grätzer, "Universal Algebra" (2nd ed.), Springer–Verlag, New York, 1979.
6. P.J. Higgins, Algebras with a scheme of operators, *Math. Nachr.* **27** (1963), 115–132.
7. G. Higman, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* (III) **2** (1952), 326–336.
8. J. Kahn, On the uniqueness of matroid representations over GF(4), *Bull. London Math. Soc.* **20** (1988), 5–10.
9. J.P.S. Kung, Strong maps, in "Theory of Matroids" (Ed. N. White), Cambridge University Press, Cambridge, 1986.
10. J.G. Oxley, On the matroids representable over GF(4), *J. Combinatorial Theory* (B) **41** (1986), 250–252.
11. N. Robertson and P.D. Seymour, graph minors – a survey, preprint.
12. G.–C. Rota, Combinatorial theory, old and new, *Proc. International Congress of Math., Nice,* 1970 (Gauthier–Villars, Paris, 1971) 229–233.
13. P.D. Seymour, Matroid representation over GF(3), *J. Combinatorial Theory* (B) **26** (1979), 159–173.
14. P.D. Seymour, The matroids with the max–flow min–cut property, *J. Combinatorial Theory* (B) **23** (1977), 189–222.

15. W.T. Tutte, A homotopy theorem for matroids II, *Trans. A.M.S.* **88** (1958), 161–174.

16. W.T. Tutte, "Graph Theory", Addison-Wesley, Menlo Park, 1984.

17. W.T. Tutte, Lectures on matroids, *J. Res. Nat. Bur. Stand.* **69B** (1965), 1–48.

18. D.J.A. Welsh, "Matroid Theory", Academic Press, London, 1976.