

# **What do Mathematicians Think Biologists Want from Supertrees? An Axiomatic Perspective**

William H. E. Day  
Port Maitland, NS B0W 2V0, Canada  
*whday@istar.ca*

12 March 2003

DIMACS Tree of Life Working Group Meeting

Biologists understand evolutionary processes well enough to have a fair idea of what they want from supertree methods to estimate and synthesize evolutionary relationships. They recognize that it would be useful to characterize or design supertree methods in terms of their properties or axioms, yet the educational systems are such that biologists may not have acquired the mathematical skills necessary to undertake such axiomatic analyses. Mathematicians can help: they like to view such problems in terms of formal models and axioms, yet the educational systems are such that their familiarity with the biological underpinnings of supertree research may be sketchy and/or simplistic. If biologists and mathematicians wish to collaborate on supertree problems, they might begin with the premise that many relevant properties are so inadequately defined, and their interrelationships so poorly understood, that usually it is impossible to obtain interesting nontrivial formal results. To address this problem, I will describe a basic framework in which agreement, consensus and supertree problems can be formulated, and in which some of the more important supertree concepts might be given precise specifications. If biologists and mathematicians find this approach relevant, we might discuss later how to extend or refine it to meet the needs of individual researchers.

0 min

# Acknowledgements

- M. Wilkinson, J. L. Thorley, D. Pisani, F.-J. Lapointe, J. O. McInerney
- F. R. McMorris
- M. A. Steel, A. W. M. Dress & S. Böcker, *Systematic Biology* 49(2):363–368(2000)

I started to prepare this talk after reading a manuscript written by Mark Wilkinson and his colleagues.

I've incorporated some ideas on aggregation models developed by Buck McMorris and me.

I have been strongly influenced by the Steel–Dress–Böcker paper, which is written for biologists and which (to my knowledge) is the only paper yet published on supertrees from an axiomatic viewpoint.

2 min

# **Working Assumptions**

## **Biologists**

say too much, imprecisely.

## **Mathematicians**

say too little, but very precisely.

Strive to occupy the middle ground: say just enough, and with reasonable precision.

4 min

# The Really, Really Important Scientific Problems of Our Time

1. What is a supertree?
2. What is a supertree problem?
3. What properties do supertree problems have?

Concerning the first problem, the biologists in this room surely understand biological supertrees and their uses as estimates of evolutionary history. As for the mathematicians, they probably don't want to know any more about supertrees than is required to construct appropriate models. So I will emphasize problems 2 and 3.

6 min

## **Why Axiomatize? (1)**

“The axiomatic method is, strictly speaking, nothing but this art of drawing up texts whose formalization is straightforward in principle. As such it is not a new invention; but its systematic use as an instrument of discovery is one of the original features of contemporary mathematics.” — Nicholas Bourbaki (1968)

In support of the axiomatic approach, I offer these inspirational readings.

...

Bourbaki was an amateur mathematician who found his vocation serving as a general in Napoleon's army. His name is used here pseudonymously.

8 min

## Why Axiomatize? (2)

“The change to an articulate mathematical symbolism well adapted to the material brought benefits of a kind and scale which . . . could not have been foreseen. Its first fruits were a series of articles in the journals, some of them dealing with fundamental aspects of the theory of committees. By axiomatizing the theory Arrow’s work had blown a sudden energy into the subject.” — Duncan Black (1991).

Black's paper is a critique of Arrow's contributions to social choice theory. Written in 1972, the year Arrow received his Nobel Prize, it was published after Black's death in 1991.

10 min

# What to Axiomatize?

## Properties of Supertree Problems

**Accuracy:** assessable, co-Pareto, independence, order invariance, Pareto, positionless, shapeless, sizeless, weightable

**Model constraints:** generality, plenary, uniqueness

**Practicality:** space, time

These properties are from Mark's manuscript and his DIMACS talk.

I will say nothing about practicality: the evaluation of time and space complexities has been well studied by computer scientists.

Some requirements can be incorporated directly in the model's specification, and so need not appear as axioms of the model.

I am primarily interested in axioms of the first type which, if satisfied by an aggregation rule, may increase our confidence in the relevance of that rule's results.

12 min

# Consensus Models

$\mathcal{X}$  = generic set of objects to be aggregated

$\mathcal{X}^k$  = set of all  $k$ -tuples or profiles of  $\mathcal{X}$

$$\mathcal{X}^* = \bigcup_{k \geq 1} \mathcal{X}^k$$

Consensus:  $C : \mathcal{X}^k \longrightarrow \mathcal{X}$

Multiconsensus:  $C : \mathcal{X}^k \longrightarrow 2^{\mathcal{X}} \setminus \{\emptyset\}$

Complete consensus:  $C : \mathcal{X}^* \longrightarrow \mathcal{X}$

Complete multiconsensus:  $C : \mathcal{X}^* \longrightarrow 2^{\mathcal{X}} \setminus \{\emptyset\}$

Since the early 1980s there has been a continuing interest in developing **consensus rules** for biological applications. Although inappropriate for investigating supertrees, consensus rules are a useful point of reference.

Invariably there is a set of voters. Each voter votes by specifying an object. The consensus rule accepts a profile of these objects and returns a unique consensus object that in some sense best represents the profile.

The basic consensus model can be varied by changing its domain and/or codomain.

14 min

# Components of Consensus Models

1. Set  $K$  of **indices** to name the voters.
2. Set  $S$  of **labels**, *e.g.*, species names, with which to describe the objects.
3. Set  $\mathcal{X}$  of **objects**, *e.g.*, hierarchies or phylogenies.
4. A **reduction** (restriction, contraction) function to exhibit the constituent parts of objects.
5. **Encoding** functions to characterize objects in meaningful ways.

Consensus models are usually specified by five components.

Invariably there is a set of voters; to name them we will use a finite set  $K$  of indices. As for the objects in  $\mathcal{X}$ , usually there is a natural set  $S$  of labels in terms of which each object can be described. Index, label and object are the initial concepts.

If we view an object as a complex entity specified in terms of labels, then we may wish to apply a reduction function to isolate parts of that object for study.

An object may have different types of relevant features, such as clusters, triads, quartets or components. Each encoding function characterizes objects in terms of such a feature.

16 min

# Analyzing Aggregation

1. Begin with concepts of **index**, **label** & **object**.
2. Define a model of *synthesis*. Specify axioms, use them to prove things.
3. Add a concept of **reduction**. Specify axioms, ...
4. Add a concept of **encoding**. Specify axioms, ...
5. Add a concept of ?. Specify axioms, ...
6. Repeat 2–5 for other models of aggregation.

We might hope that such components will occur in any aggregation model that synthesizes small objects into a single large object. Here is a plausible strategy for designing such models.

18 min

## Initial Concepts

$K = \{1, \dots, k\}$ , a set of things called *indices*

$S = \{s_1, \dots, s_n\}$ , a set of things called *labels*

$(\forall X \subseteq S)(\mathcal{X}_X = \text{a set of things called } \textit{objects}, \text{ each defined in terms of each and every label of } X)$

$$(\forall X \subseteq S)(\mathcal{X}_{[X]} = \bigcup_{Y \subseteq X} \mathcal{X}_Y)$$

$\mathcal{X} = \mathcal{H}_S$ , *i.e.*, hierarchies with exactly  $n$  leaf labels

$\mathcal{X} = \mathcal{H}_{[S]}$ , *i.e.*, hierarchies with at most  $n$  leaf labels

To begin the design process, here are the three basic components for specifying models of aggregation.

There is an important distinction between  $\mathcal{X}_X$  and  $\mathcal{X}_{[X]}$ , the two basic sets of objects: an object of  $\mathcal{X}_X$  must have the label set  $X$ , while an object of  $\mathcal{X}_{[X]}$  may have any label set that is a subset of  $X$ . Clearly  $\mathcal{X}_X \subseteq \mathcal{X}_{[X]}$ .

20 min

# Models of Aggregation

For  $C$  a partial function,

Agreement:  $C : \mathcal{X}_S^k \longrightarrow \mathcal{X}_{[S]}$

Consensus:  $C : \mathcal{X}_S^k \longrightarrow \mathcal{X}_S$

Synthesis:  $C : \mathcal{X}_{[S]}^k \longrightarrow \mathcal{X}_S$

Multisynthesis:  $C : \mathcal{X}_{[S]}^k \longrightarrow 2^{\mathcal{X}_S} \setminus \{\emptyset\}$

Complete synthesis:  $C : \mathcal{X}_{[S]}^* \longrightarrow \mathcal{X}_S$

Complete multisynthesis:  $C : \mathcal{X}_{[S]}^* \longrightarrow 2^{\mathcal{X}_S} \setminus \{\emptyset\}$

With the three basic concepts we can specify three basic types of aggregation model: agreement, consensus and synthesis.

Just as consensus models had four variants, so do synthesis models; but today I will only discuss the basic synthesis model.

22 min

## Conventions

( $\forall$  functions  $f, g$ ),  $fgT$  means  $f(g(T))$ .

( $\forall P = (T_1, \dots, T_k) \in \mathcal{X}_{[S]}^k$ ),  $P$  is called a *profile*.

( $\forall T \in \mathcal{X}_{[S]}$ ),  $(T)_k = (T_1, \dots, T_k) = (T, \dots, T) \in \mathcal{X}_{[S]}^k$   
is a *constant* profile.

$\ell : \mathcal{X}_{[S]} \longrightarrow 2^S$  displays an object's labels:

( $\forall T \in \mathcal{X}_{[S]}$ )( $\ell T =$  the set of labels for  $T$ )

Several conventions make the following developments easier to grasp.

24 min

## Collective Rationality (*ColRat*)

$(\forall P \in \mathcal{X}_{[S]}^k)(CP \text{ is well-defined})$

Our first axiom, collective rationality, requires that the synthesis rule be well-defined, *i.e.*, that it return for each profile a unique object as that profile's synthesized result.

Although collective rationality is usually assumed for consensus rules, it may be undesirable for synthesis rules. For example, if for a given profile there is no significant overlap among the objects, then how could any synthesis rule return a meaningful synthesized result?

## Index Concept

$\sigma : K \longrightarrow K$  permutes elements of  $K = \{1, \dots, k\}$

$\sigma : \mathcal{X}_{[S]}^k \longrightarrow \mathcal{X}_{[S]}^k$  permutes object indices in profiles:

$$(\forall P = (T_1, \dots, T_k) \in \mathcal{X}_{[S]}^k) [\sigma P = (T_{\sigma 1}, \dots, T_{\sigma k})]$$

## Index Symmetry (*I-Sym*)

$$(\forall P \in \mathcal{X}_{[S]}^k) (\forall K\text{-permutations } \sigma) (CP = C\sigma P)$$

*i.e.*,  $(\forall K\text{-permutations } \sigma) (C = C\sigma)$

Our second axiom concerns the index concept.  $\sigma$  is a permutation function on  $K$ , *i.e.*, it is a bijection and so is one-one and onto. It is straightforward to use  $\sigma$  to permute objects in a profile by permuting the object indices. Although both functions are named  $\sigma$ , context always shows which  $\sigma$  is intended.

The index symmetry axiom requires that a synthesized result be invariant under every permutation of the objects in every profile.

# Label Concept

$\phi : S \longrightarrow S$  permutes labels in  $S$

$\phi : \mathcal{X}_{[S]} \longrightarrow \mathcal{X}_{[S]}$  permutes labels of objects:

$(\forall T \in \mathcal{X}_S)(\phi T = \text{object obtained}$

by using  $\phi$  to permute the labels of  $T$ )

$\phi : \mathcal{X}_{[S]}^k \longrightarrow \mathcal{X}_{[S]}^k$  permutes labels of profile objects:

$(\forall P \in \mathcal{X}_{[S]}^k)[\phi P = (\phi T_1, \dots, \phi T_k)]$

Our third axiom concerns the label concept.  $\phi$  is a permutation function on  $S$ , *i.e.*, it is a bijection and so is one-one and onto. It is straightforward to use  $\phi$  to permute labels in a single object, and by extension to permute labels in every object in a profile. Although all three functions are named  $\phi$ , context always shows which  $\phi$  is intended.

30 min

## Label Consistency (*L-Cons*)

$$(\forall P \in \mathcal{X}_{[S]}^k)(\forall S\text{-permutations } \phi)(C\phi P = \phi C P)$$

*i.e.*,  $(\forall S\text{-permutations } \phi)(C\phi = \phi C)$

The label consistency axiom requires that a synthesized result be invariant under every permutation of the object labels.

For each profile  $P$  and  $S$ -permutation  $\phi$ : if we use  $\phi$  to relabel  $P$  and then apply  $C$  to that modified profile, the synthesized result is equal to the result of applying  $C$  to  $P$ , but with its labels then permuted by  $\phi$ .

“If we rename all the species, and then apply the method to the new input trees, the output tree is simply the old output tree, but with the species renamed accordingly.”

— Steel et al. (2000, p. 364)

32 min

## Object Pareto (*O-Par*)

$$(\forall T \in \mathcal{X}_S)(\forall P \in \mathcal{X}_{[S]}^k) \\ [(\forall i \in K)(T = T_i) \implies T = CP]$$

## Object Autonomy (*O-Aut*)

$$(\forall T \in \mathcal{X}_S)(\exists P \in \mathcal{X}_{[S]}^k)(T = CP)$$

Four axioms concern objects.

The first is a form of optimality: for each object having  $S$  as its set of labels and for each profile, if that object is at every position of the profile, then the synthesis rule must return that object as the synthesized result.

The second is a form of autonomy: every object of the synthesis rule's codomain must be in its range. That is, for each object having  $S$  as its set of labels, a profile must exist for which that object is the synthesized result.

## Object Dictatorship (*O-Dict*)

$$(\exists j \in K)(\forall P \in \mathcal{X}_{[S]}^k)(T_j = CP)$$

## Object Co-Pareto (*O-CoP*)

$$(\forall P \in \mathcal{X}_{[S]}^k)(\ell CP \subseteq \bigcup_{i \in K} \ell T_i)$$

The third specifies the behavior of an object dictator. There exists an index  $j$  such that, for every profile, the object at that profile's  $j^{\text{th}}$  position must be returned as the synthesized result.

Object dictators are impossible if index symmetry holds.

The fourth specifies a criterion of parsimony. Every label of a synthesized result must be in at least one of the objects in any profile yielding that synthesized result.

Surely every reasonable supertree rule is object Co-Pareto.

## Reduction Concept

$(\forall X \subseteq S)$ ,  $\xi_X : \mathcal{X}_{[S]} \longrightarrow \mathcal{X}_{[X]}$  reduces objects on  $S$  to objects on  $X$ :  $(\forall T \in \mathcal{X}_{[S]})$ ,  $\xi_X(T)$  is obtained by suppressing in  $T$  the structure associated with  $S \setminus X$ .

$(\forall X \subseteq S)$ ,  $\xi_X : \mathcal{X}_{[S]}^k \longrightarrow \mathcal{X}_{[X]}^k$  reduces profiles on subsets of  $S$  to profiles on subsets of  $X$ :

$$(\forall P \in \mathcal{X}_{[S]}^k)[\xi_X P = (\xi_X T_1, \dots, \xi_X T_k)]$$

$$(\forall P, P' \in \mathcal{X}_{[S]}^k)[P = P' \iff (\forall i \in K)(T_i = T'_i)]$$

Reduction is like an X-ray machine that penetrates the surfaces of objects to reveal their hidden structure.

For any subset  $X$  of  $S$ , and for any object  $T$  on  $S$ , the reduction function  $\xi_X$  enables us to study a corresponding object on  $X$ . Thus if the original object is a graph  $G$  with vertex set  $S$ , then the reduction function for  $X$  might return the subgraph of  $G$  that is induced by  $X$ .

It is easy to extend the reduction function so that it operates on profiles of objects, rather than single objects.

Two profiles are called equal if the corresponding objects at every index position are equal.

## Reduction Commutativity (*R-Com*)

$$(\forall P \in \mathcal{X}_{[S]}^k)(\forall X, Y \subseteq S)(\xi_X \xi_Y P = \xi_Y \xi_X P)$$

*i.e.*,  $(\forall X, Y \subseteq S)(\xi_X \xi_Y = \xi_Y \xi_X)$

What makes a reasonable reduction function?

Suppose we have a large tree  $T$  with labeled leaves, and we wish to reduce  $T$  to a smaller tree by pruning the three leaves labeled  $x$ ,  $y$  and  $z$ . We expect the result to be insensitive to the order in which those leaves are pruned.

Reduction commutativity ensures that the result of applying several reductions is independent of the order in which they are applied.

This axiom is unusual: it does not mention the synthesis rule.

40 min

## Reduction Independence (*R-Inde*)

$$(\forall P, P' \in \mathcal{X}_{[S]}^k)(\forall X \subseteq S)$$

$$(\xi_X P = \xi_X P' \implies \xi_X CP = \xi_X CP')$$

## Reduction Consistency (*R-Cons*)

$$(\forall P \in \mathcal{X}_{[S]}^k)(\forall X \subseteq S)(C\xi_X P = \xi_X CP)$$

*i.e.*,

$$(\forall X \subseteq S)(C\xi_X = \xi_X C)$$

Reduction independence is the concept of Arrowian independence. Consider any two profiles of objects, and consider any subset  $X$  of  $S$ . If we reduce the profiles to  $X$  and find that the reduced profiles are equal, then Arrowian independence requires that the corresponding synthesized objects, when reduced to  $X$ , must also be equal.

Reduction consistency concerns whether the order in which we apply the reduction and synthesis functions is critical. Consider any profile  $P$  of objects, and consider any subset  $X$  of  $S$ . Reduction consistency requires that the reduction to  $X$  of an object synthesized from  $P$  be equal to the object synthesized from  $P$  after it has been reduced to  $X$ .

42 min

# Reduction Confusion

$$R\text{-Inde} = R\text{-Cons} (!)$$

“Borrowing Arrow’s terminology we shall say that in the kind of cases described above the minimax regret solution is ‘dependent upon irrelevant alternatives.’”

— Radner & Marschak (1954)

“Arrow (1951, pp. 26–27) surrounds his definition of [*R-Inde*] with two examples of [*R-Cons*] and one of [*R-Inde*].” — McLean (1995)

In the 1950s confusion arose on the meaning of independence because the concepts of reduction independence and reduction consistency were confounded.

For me there is only one generic concept of independence, and it is Arrow's. If another concept does not look and feel like Arrowian independence, it should not be called independence. Accept no substitutes: for independence, Arrow's is the only game in town!

44 min

## Display

$$(\forall T_1, T_2 \in \mathcal{X}_{[S]})$$

$$[T_1 \text{ displays } T_2 \text{ if } (X = \ell T_2 \text{ and } \xi_X T_1 = T_2)]$$

$$(\forall T \in \mathcal{X}_{[S]})(\forall P \in \mathcal{X}_{[S]}^k)$$

$$[T \text{ displays } P \text{ if } (\forall i \in K)(T \text{ displays } T_i)]$$

## Reduction Display (*R-Disp*)

$$(\forall P \in \mathcal{X}_{[S]}^k)$$

$$[(\exists T \in \mathcal{X}_{[S]})(T \text{ displays } P) \implies (CP \text{ displays } P)]$$

The display concept extends the reduction function so that it applies to a profile of objects.

“Display” is a convenient synonym for “reduce to:” for all objects  $T_1$  and  $T_2$ ,  $T_1$  displays  $T_2$  if and only if  $T_1$  reduces to  $T_2$ .

For each profile  $P$ , an object  $T$  displays  $P$  if  $T$  displays every  $T_i$  in  $P$ . Notice that phylogenetic trees in a profile  $P$  are compatible if and only if some tree displays  $P$ .

Reduction display requires, for every profile  $P$ , that the synthesized object display  $P$  if some object displays  $P$ .

## Encoding Concept

$E_S =$  a set of *atoms* defined by the labels in  $S$ .

Examples of atoms: clusters, triads, quartets.

$e : \mathcal{X}_{[S]} \longrightarrow 2^{E_S}$  encodes objects by sets of atoms:

$(\forall T \in \mathcal{X}_{[S]})(eT \text{ is a set of atoms characterizing } T)$

Each object has had associated with it only a set of labels. Now we will use an encoding scheme to characterize each object in terms of a set of elementary atomic structures.

This is a natural step since hierarchies or rooted phylogenetic trees can each be characterized by a set of clusters or by a set of triads, and phylogenies or unrooted phylogenetic trees can each be characterized by a set of quartets.

More than one encoding might be relevant to any given synthesis problem.

48 min

## Encoding Pareto (*E-Par*)

$$(\forall x \in E_S)(\forall P \in \mathcal{X}_{[S]}^k)[(\forall i \in K)(x \in eT_i) \implies x \in eCP]$$

## Encoding Autonomy (*E-Aut*)

$$(\forall x \in E_S)(\exists P \in \mathcal{X}_{[S]}^k)(x \in eCP)$$

Four axioms concern objects encoded in this way. The axioms are strictly analogous to those we discussed previously.

These two are analogous to object Pareto and object autonomy.

50 min

## Encoding Dictatorship (*E-Dict*)

$$(\exists j \in K)(\forall P \in \mathcal{X}_{[S]}^k)(eT_j \subseteq eCP)$$

## Encoding Co-Pareto (*E-CoP*)

$$(\forall P \in \mathcal{X}_{[S]}^k)(eCP \subseteq \bigcup_{i \in K} eT_i)$$

These two are analogous to object dictatorship and object Co-Pareto.

52 min

## Concepts & their Axioms

Slide	Concept		Axiom	Axiom
13	—		ColRat	
14	Index	●★	I-Sym	
16	Label	●★	L-Cons	
17	Object		O-Par	O-Aut
18	Object		O-Dict	O-CoP
20	Reduction		R-Com	
21	Reduction	○	R-Inde	★ R-Cons
23	Reduction	●	R-Disp	
25	Encoding	●○★	E-Par	E-Aut
26	Encoding	○	E-Dict	★ E-CoP

○ Arroviaian. ● Steel et al. (2000). ★ Wilkinson et al.

This table of contents shows where the axioms are defined. It reminds us that this study is based on and driven by five basic concepts. It flags several axioms that have been used by other investigators. Notice that the three investigators mentioned chose to focus on three distinct, important aspects of the reduction concept.

We have a long way to go on the quest to specify and investigate axioms of the types mentioned by Mark Wilkinson and his colleagues. The approach I have outlined has limitations; the project to formalize properties relevant to biologists may be infeasible.

54 min

# Complete Multisynthesis Model

$$C : \mathcal{X}_{[S]}^* \longrightarrow 2^{\mathcal{X}_S} \setminus \{\emptyset\}$$

For studying complete multisynthesis **median** rules.

## Problems

0. Do the axioms and results for complete multi-consensus **median** rules generalize to this model?

Here are six problems of varying degrees of relevance or significance. To motivate this one, recall that for the consensus model a great deal is now known concerning multiconsensus median rules.

0. I hope so. Why hasn't someone already done this?

56 min

## More Problems

1. How to combat terminological confusion?
2. Where was the biology in this talk?
3. Are Arrowian impossibility results possible in the synthesis model?
4. How to accommodate sizeless, shapeless, *etc.*, in the synthesis model?
5. How to formulate synthesis models from a lattice-theoretic point of view?

1. Has this talk helped? Or is it part of the problem?
2. In the undefined terms such as object, reduction, encoding.
3. I hope so. Why hasn't someone already done this?
4. Even defining such concepts may be problematic. For each such concept, is there a particular encoding that provides a natural setting in which the concept can be studied?
5. This may require a substantial change in point of view, one that biologists may find difficult to make.

58 min