Wholes and Parts in General Systems Methodology*

Martin Zwick
Systems Science Ph.D. Program
Portland State University
zwickm@pdx.edu http://www.sysc.pdx.edu/Faculty/Zwick

Abstract

Reconstructability analysis (RA) decomposes wholes, namely data in the form either of set-theoretic relations or multivariate probability distributions, into parts, namely relations or distributions involving subsets of variables. Data is modeled and compressed by variable-based decomposition, by more general state-based decomposition, or by the use of latent variables. Models, which specify the interdependencies among the variables, are selected to minimize error and complexity.

key words: wholes, parts, system, relation, constraint, structure, information, uncertainty, entropy, complexity, latent variables, state-based modeling, discrete multivariate modeling, reconstructability analysis, information theory, set-theoretic modeling, log-linear methods, systems theory

I. Reconstructability Analysis

In general systems methodology, the decomposition of wholes into parts and the composition of parts into wholes is called *reconstructability analysis* (Klir, 1985). RA derives from the early work of Ashby (1964), and was developed by Broekstra, Cavallo, Conant, Jones, Klir, Krippendorff, and other researchers (see the citation in the reference list below of an RA bibliography in the International Journal of General Systems). Here, a whole is a *relation* which is a *constraint* among a set of variables. The parts, which define the *structure* of the whole, are relations among subsets of the variables, subsets which may be either disjoint or overlapping. For a fully decomposable whole, the parts are simply the variables. For example, for variables, A, B, and C, the whole is the triadic relation ABC. The parts might be the dyadic relations, AB and BC, in which case the system has structure AB:BC. (A colon is used to separate the parts. The whole, when maximally decomposed, yields A:B:C.) Data defines a whole and RA reveals its parts. The parts taken together are specified by fewer numbers (*parameters*) than the whole, i.e., decomposition accomplishes "compression." An RA-generated structure is a *model* of the data.

The variables in RA are nominal (categorical, qualitative). Thus RA might be used, for example, in studying the interactions between biological characters or genes, and their relations to environmental conditions, since characters, genes, and conditions may be represented as nominal variables. RA has considerable value also for analyzing quantitative variables linked by unknown nonlinear relations. (For linear relations, standard methods are superior.) Quantitative variables are accommodated by "binning" their values into discrete states which are unordered, and binning can be done within the framework of fuzzy set theory (Zadeh, 1965) or by clustering techniques. RA can analyze not only static relations, but also dynamic relations: in ABC, variables might be state(t-2), state(t-1), and state(t). Relations can be deterministic or stochastic.

***The Character Concept in Evolutionary Biology**, Gunter Wagner, Ed., Academic Press, 2001.

There are two main formalisms used in RA to define a relation (Conant, 1981; Klir, 1985; Krippendorff, 1986). These are here called the set-theoretic and the information-theoretic. Set-theoretically, a relation is a subset of a cartesian product: the combinations of possible variable values which are actually observed. Information-theoretically, a relation is a multivariate probability distribution. More precise labels for these formalisms are "crisp possibilistic" and "probabilistic," but the set- and information-theoretic labels are used here because they are more familiar and because this nomenclature correctly suggests parallelisms in the formalisms. The set- and information-theoretic perspectives are central to the general systems literature and can now also be viewed as components of a generalized information theory currently being developed (Klir & Wierman, 1998).

The information-theoretic aspect of RA overlaps with log-linear modeling (Bishop et al, 1978; Knoke & Burke, 1980) which is widely used for analyzing nominal data in the social sciences. Log-linear modeling was developed in the same period as RA , but the general systems methodology literature is broader. Log-linear modeling is statistical, while reconstructability analysis also includes well-developed non-statistical aspects, e.g., in its set-theoretic formalism, lattice explorations, advanced computational algorithms, and analysis of fuzzy distributions. It also differs from log-linear modeling in its extensive use of uncertainty measures, in its innovation of "state-based" modeling, and in its acceptance of the challenge of modeling data on many variables. On the other hand, the log-linear literature is very advanced statistically, while statistical considerations are sometimes absent in the systems literature where their presence would be desirable. Latent variable techniques, which are the nominal data analog of factor analysis, are well developed on the log-linear side, while on the systems side they are available mainly set-theoretically. Where information-theoretic and log-linear methods overlap, they are equivalent; compare, e.g., Knoke & Burke (1980) and Krippendorff (1986).

In Section II, RA is illustrated with simple examples of both set- and information-theoretic analyses, so the reader can easily grasp what is fed into RA and what it yields in return. To explain RA strictly in these input-output terms requires only a minimal mathematical description. The theory is then presented in Sections III and IV, which explain what relations and structures are. The practice of RA is described in Section V, i.e., the "black box" is opened and the analytical methods used to obtain the results of Section II are described. Section VI closes the paper with remarks on the current state of RA methodology.

II. Examples

Table 1 illustrates with two examples: (a) set-theoretic and (b) information-theoretic. In both examples, variables A, B, and C are dichotomous (binary). If variables had three states (values), e.g., $\{A_0,A_1,A_2\}$, the order of the states would be arbitrary because the variables are nominal.

The input to RA is shown on the left of the table. In (a) the data are the combinations (tuples, binary strings) observed for the variables -- five tuples of the possible eight -- without regard to how frequently they are observed. In (b) the A, B, and C values are labeled with 0 and 1 subscripts (instead of being taken as 0 or 1); all combinations are observed, but with the frequencies given in the contingency table.

Table 1. System as data and model (two examples: (a) & (b))

data: one trivariate relation ABC — model: two bivariate relations AB : BC

(a) ABC = {000,010,011,110,111} —RA→ AB : BC ={00,01,11}:{00,10,11}

(b)

	B_0		B_1	
	C_0	C_1	C_0	C_1
A_0	143	253	77	182
A_1	227	411	46	139

RA →

	B_0	B_1
A_0	396	259
A_1	638	185

	B_0	B_1
C_0	370	123
C_1	664	321

The task of RA is to decompose the ABC relation into parts in such a way that the model, simpler than the data, still adequately agrees with the data. The information-theoretic analysis is statistical; the set-theoretic analysis is non-statistical. Partial RA results are shown on the right side of Table 1. Both data (a) and (b) are decomposable into structure AB:BC. The set-theoretic model consists of a set of AB tuples and a set of BC tuples. Taken together, these are *equivalent* to the ABC data (the 5 tuples) shown on the left of part (a) of the table. Model and data agree exactly, i.e., with no error.

The information-theoretic model analogously consists of the two 2-variable contingency tables, AB and BC, as shown on the right of Table 1(b). Taken together, these are equivalent to a 3-variable table which *approximates* the data on the left. Data and model, expressed as probabilities and not frequencies, are shown in Table 2. The data (the p distribution) and the model (the q distribution) do not agree exactly, but the error is statistically acceptable. Error is loss of constraint or equivalently, loss of information.

Table 2. ABC (data) and $ABC_{AB:BC}$ (model) (data from Table 1(b))

	B_0		B_1	
	C_0	C_1	C_0	C_1
A_0	.097	.171	.052	.123
A_1	.154	.278	.031	.094

ABC data: p(A,B,C)

	B_0		B_1	
	C_0	C_1	C_0	C_1
A_0	.096	.172	.049	.127
A_1	.154	.277	.035	.091

AB:BC model: $q_{AB:BC}(A,B,C)$

Detailed explanations of these set- and information-theoretic analyses is given in Sections IV and V. A complete RA does not merely yield a single model which adequately fits the data. It can determine how well *all possible* models fit the data; or, if the number of possible models is too large to evaluate exhaustively, RA searches through a promising subset of them. Table 3 summarizes the analysis of all models for the set-theoretic example of Table 1(a). There are 9 possible models, descending from the most complex (the triadic relation, ABC, i.e., the data, also called the "saturated" model) at the top to the simplest (three "monadic" relations, A:B:C, also called the "independence" model) at the bottom.

After each model in parenthesis is indicated the number of ABC tuples which the model specifies. The top model, ABC, is the data itself which contains 5 ABC tuples. Models AB:AC:BC and AB:BC also specify the same 5 tuples. These two models fit the data without error, but since AB:BC is less complex, it is preferred. Other models specify either 6 or 8 tuples, that is, they predict additional combinations which are not observed in the data and are thus in error. There is less constraint represented in these models than is in the data. Models which predict 8 tuples predict all possible combinations of A, B, and C, and exhibit no constraint at all.

Table 3. Reconstruction of the set-theoretic data of Table 1(a). in parentheses: (number of tuples in model)

	ABC (5)	
	AB:AC:BC (5)	
AB:AC (6)	**AB:BC (5)**	BC:AC (6)
AB:C (6)	AC:B (8)	BC:A (8)
	A:B:C (8)	

Table 4 summarizes the analysis of all models for the information-theoretic example of Table 1(b). After each model three numbers are given in parentheses:

- Information^{norm}, the information (constraint represented) in the model, which ranges from 0 (no information) to 1 (complete information),
- α , the probability of making an error (called a Type I error) if one rejects the identity of the model with the data, and
- df, the degrees of freedom (complexity) of the model.

At the top of the list, the "saturated" model ABC has complete information. The probability of making an error in rejecting its identity with the data is 1, since it *is* the data. Its degrees of freedom, i.e., the number of probabilities needed to specify the 3-variable table, is 7. At the bottom of the list, the "independence" model A:B:C has no information. The probability of making an error in saying it is different from the data is 0. It needs 3 probability values for its specification. In between top and bottom models are all less-than-total decompositions. A good model is one which has high information content and small df (complexity). It is also a model which has high α. (The desirability of a high α is atypical of most statistical analyses, but if we compared a model not to the saturated model at the top but to the independence model at the bottom, the normal preference for low α values would obtain.) The best model is AB:BC, but the analysis tells us more than this. It tells us how well all possible structures model the data.

Table 4. Reconstruction of the information-theoretic data of Table 1(b). in parentheses: (Information^{norm} , α, df)

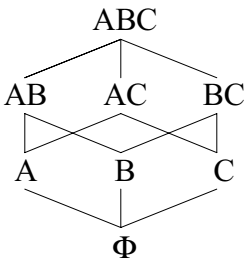
	ABC (1.,1., 7)	
	AB:AC:BC (.987, .382, 6)	
AB:AC (.827, .005,5)	**AB:BC (.978. .518, 5)**	BC:AC (.153, .000, 5)
AB:C (.826, .014, 4)	AC:B (.000, .000, 4)	BC:A (.152, .000, 4)
	A:B:C (0., .000, 3)	

III. Relations

A. Lattice of Relations

For a system of two variables, there are only two possible structures: AB and A:B. For three variables, the relations which could exist are arrayed in the "lattice of relations" shown in Figure 1 (Krippendorff, 1986). The number of variables in a relation is its *ordinality*. In this framework, relations can have arbitrary ordinality. By contrast, the conventional graph-theoretic representation which depicts systems in terms of nodes and links connecting nodes normally restricts ordinality to two. A relation obtained from a higher-ordinality relation by ignoring one or more variables is "embedded in" and "a projection of" the higher-ordinality relation, e.g., AC is embedded in ABC. Although a relation viewed as a *constraint* usually presumes the existence of at least two variables, it is useful to include "monadic relations," A, B, and C, and, for completeness, the "null relation," Φ.

Figure 1. Lattice of relations for a 3-variable system.



Relations are *directed* if a variable acts on another, or an event causes another, or a discrimination is made between "generating" or "generated" variables. "Generating" vs. "generated" is used instead of the more familiar "independent" (IV) vs. "dependent" (DV) (variables) to reserve the word "independent" for its connotations of "mutually independent" or "independent of." Relations are *neutral* if directionality cannot be or is not indicated. Directed relation AB is distinguished from directed relation BA. Neutral relations are static, but directed relations can be static (e.g., nontemporal input-output pairs) or explicitly dynamic. Directed relations may be deterministic or stochastic.

B. Definition of Relation

Set-theoretically (Conant, 1981; Klir, 1985), $AB \subseteq A \otimes B$: a relation AB is a subset of the Cartesian product of sets, $A \otimes B$, where A and B signify also the sets of states the variables take on. The Cartesian product -- call it H (for "heap") -- is the set of all possible pairs of values of the two variables. A relation is a constraint which reduces the possible to the actual, i.e., AB reduces H to a smaller set of pairs $\{(A_i,B_j)\}$ which in fact occur. (If the pairs are allowed *partial* membership in set AB, the relation is fuzzy.) H can be written also as the independence model A:B, where the colon means mutually independent. The number of pairs (in general, "tuples") in AB is called its cardinality and is written |AB|.

Referring again to Figure 1, projection into lower-ordinality relations ignores the projected variables. For example, AB is obtained from ABC by ignoring all C_k in $\{(A_i,B_j,C_k)\}$; the

monadic relation (variable) A is obtained by ignoring both B_j and C_k. For the ABC relations shown on the left of Table 1(a), the AB and BC projections are shown on the right.

Information-theoretically (Klir, 1985; Krippendorff, 1986), a relation is a multivariate probability distribution, i.e., AB is the set $\{p(A_i,B_j)\}$, for convenience sometimes written simply as p(A,B). Technically "relation" means set-theoretic relation, but the word is here used more broadly. The number of probability values needed to specify the relation is its "degrees of freedom," df(AB) = |AB|-1, where one subtracts 1 from the number of (A_i,B_j) pairs since probabilities must add to 1. For example, if A and B are dichotomous, 3 probability values define the contingency table for relation AB although the table itself has 4 entries. Constraint is the deviation of the distribution from a reference distribution, usually A:B, but occasionally Φ, here the uniform distribution. The A:B distribution is $q_{A:B}(A,B) = p(A)\ p(B)$, where $q_{A:B}$ plays the same role as the Cartesian product H. "q" denotes a *calculated* distribution and "p" denotes data, i.e., an *observed* distribution and its projections. Projection into lower-ordinality relations sums over projected variables, e.g., $p(A,B) = \sum_k p(A,B,C_k)$. Again, in Table 1(b), the ABC distribution shown on the left has AB and BC projections shown on the right.

Relations are characterized by *uncertainty*, a measure of variety or dispersion. Uncertainty is the nominal variable analogue of variance. Set-theoretic uncertainty is the Hartley entropy, $U = \log_2 |AB|$, i.e., the log of the number of pair values. Information-theoretic uncertainty is the Shannon entropy, $U = -\sum p(A,B) \log_2 p(A,B)$. ("Uncertainty" is preferable to "entropy" as "entropy" evokes an association with the Second Law of Thermodynamics and is best reserved for physical systems.) Uncertainty measures can be defined for generalizations of the set- and information-theoretic formalisms. There is not just one concept of uncertainty and a complex relationship exists between its uses in different formalisms (Klir & Wierman, 1998).

A relation is a whole whose parts are projections. The uncertainty of a whole, U(AB), is less than or equal to the uncertainty of its parts, U(A:B) = U(A) + U(B). The strength of constraint of AB in both formalisms is the uncertainty reduction, U(A:B)-U(AB); information-theoretically, this is also called "transmission" ("mutual information"), T(A:B). This is the *gain* of constraint in AB relative to A:B or equivalently the *loss* of constraint in A:B relative to AB. The uniform distribution, Φ, might serve as an alternative to A:B as a reference condition.

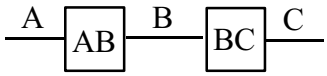
IV. Structures

A. Lattice of Structures

A two-variable system can have only one (undirected) relation, but with three or more variables, systems can have multiple relations, i.e., *structure*. A structure is an unordered set of relations none of which is a projection of another. In representations of systems where relations are strictly dyadic (involve only two variables), variables are often shown as nodes (or circles) and relations as lines or arrows connecting nodes. In the present

framework, relations may involve an arbitrary number of variables, and to facilitate focusing on relations rather than variables, structures are represented with relations as boxes and variables as lines, as shown in Figure 2 (Klir, 1985, Krippendorff, 1986). One can add directedness by changing lines into arrows.

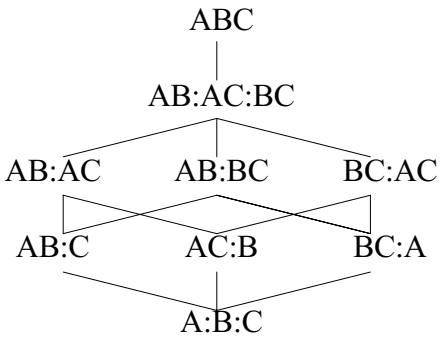
Figure 2. Specific structure AB:BC



Relations may overlap in their variables as in AB:BC or may be disjoint, as in AB:C, or completely disjoint, as in the "heap," A:B:C.

The lattice of possible (undirected) structures with three variables is shown in Figure 3 (Krippendorff, 1986). Table 3 and Table 4 simply list the possible structures; this figure shows how they are related to one another by descent.

Figure 3. Lattice of Specific Structures for 3-variable system (undirected relations)



For three variables there are nine specific structures as show in Figure 3 but only five *general* structures: (1) XYZ, (2) XY:XZ:YZ, (3) XY:YZ, (4) XY:Z, and (5) X:Y:Z. (NB: the "general" vs. "specific" nomenclature here differs from Klir's (1985).) Specific structures are obtained from general structures by permuting the assignments of specific variables, A, B, or C to the generic X, Y, or Z. For example general structure (4) XY:Z subsumes AB:C, AC:B, and BC:A. For three variables, the lattice of general structures is small (5 general structures), but for four variables (add the variable W), there are 20 general structures (Figure 4) (Klir, 1985, Krippendorff, 1986). If variables are all dichotomous (binary), the degrees of freedom of the structures range from 15 at the top to 4 at the bottom and decrease by 1 at every level. The figure also shows the acyclic structures, which are indicated with boxes in bold (10 of the 20). For four variables, there are 114 specific structures.

For directed systems the lattice is often simpler. Say W is generated from X, Y, and Z. It is assumed for directed systems that we are uninterested in relations among the generating variables, so structures always include an XYZ component which subsumes all such relations, and decompositions of this XYZ component are never considered. (If one wants to know about relations among the generating variables, the system is treated as neutral.) All the other relations in a structure necessarily involve the generated variable(s). For three

generating variables and one generated variable, only 9 of the 20 structures of Figure 4 apply: the top 6 and the next leftmost 3 (indicated in the figure with the generated variable -- a line or connected lines uninterrupted by a box -- in bold.) 4 of these 9 directed structures are also acyclic:(1) XYZ:W, (2) XYZ:XW, (3) XYZ: XYW, and (4) XYZW. These are the simplest decompositions of W's dependence on X, Y, and Z; they specify W as either independent of or dependent on 1, 2, or 3 of the generating variables. Note that XYZ:W (indicated on Figure 4 by a *), and *not* X:Y:Z:W, is the bottom of this lattice of directed structures. Also that a structure like XYZ:XW:YW:ZW is *different* from XYZW, even though in both cases W depends upon X, Y, and Z (Zwick, 1996).

Combinatorial possibilities rapidly expand for five or more variables (Table 5). Although exhaustive consideration of all structures becomes prohibitive at around 6 variables, intelligent heuristics can accommodate many more variables (Klir, 1985, Krippendorff, 1986, Conant, 1988). The lattice can be pruned as a search procedure descends or ascends so that consideration is restricted only to promising candidates. Or, the search done first roughly between groups of structures and then finely within these groups (Klir, 1985).

Table 5. Numbers of structures. Only the bottom line is for directed structures.

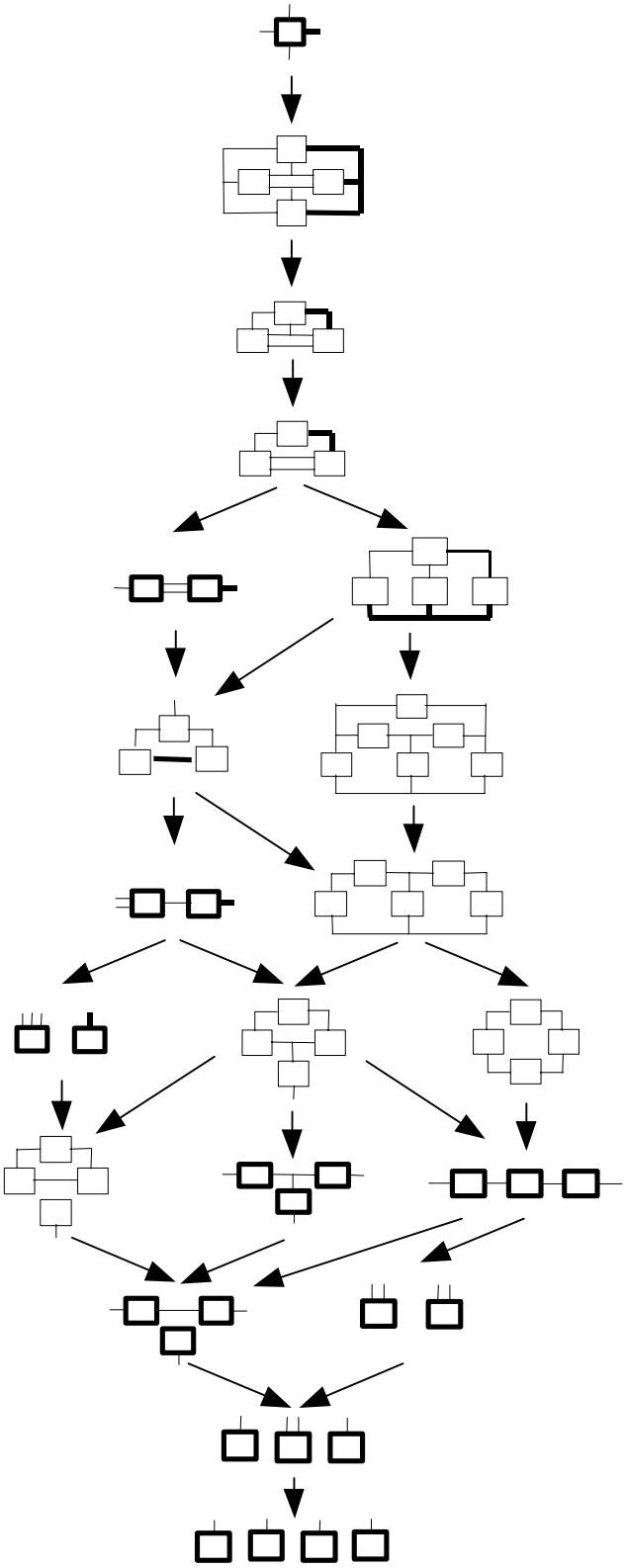
# variables	3	4	5	6
# general structures	5	20	180	16,143
# specific structures	9	114	6,894	7,785,062
with 1 generated var	5	19	167	7,580

B. Cycles, Paths, Latent Variables, State Models

The possibility of (non-trivial) *cyclicity* emerges with 3 or more variables, as in AB:BC:CA. If the relations are directed, the structure's cyclicity is directed, showing *feedback*. Methodologically cyclicity is a source of complications (Krippendorff, 1986). *Mediation* can occur by overlapping relations, and if relations are dyadic and directed (a digraph), the structure has *paths*. In AB:BC, for example, B mediates between A and C; if AB:BC is directed, B transmits the *indirect* effect of A on C via the path $A \rightarrow B \rightarrow C$. (This is different from the effect of B on AC within an undecomposed triadic relation ABC.) In directed structure AB:BC:AC, there is also an $A \rightarrow C$ path which transmits the *direct* effect of A on C. This is the nominal analog of path analysis (Davis, 1985).

Consider directed structure BA:BC. B might be a *higher-level* latent variable (or "construct") which "chunks" together (Simon, 1981) A and C. In factor analysis B would be called a "common factor." Latent class analysis does a similar analysis for nominal variables. For example, a relation AC might be explained by a latent variable B and a posited relation ABC, which subsumes AC and is decomposable into BA:BC. The latent class procedure does not actually distinguish between the factor analytic and path analytic situations. An AC relation with latent B and inferred ABC describes also directed structures where B is prior to A and C, i.e., $A \leftarrow B \rightarrow C$, or where B is intermediate between them, i.e., $A \rightarrow B \rightarrow C$.

Figure 4. Lattice of General Structures (4-variable system). A box is a relation; a line, with branches, uninterrupted by a box, is a variable. Arrows indicate decomposition. The top structure is XYZW; the fourth down is XW:XYZ:WYZ; the bottom is X:Y:Z:W. Generated variable W is shown in bold for the 9 structures of directed 4-variable systems. The 10 acyclic structures have all relations shown in bold.



The information-theoretic framework and its log-linear equivalent (Hagenaars, 1993) thus generalizes to nominal data the more restricted methods of path analysis, factor analysis, and covariance structure modeling (Long, 1983) which apply only to linear relations. Latent variable methods apply also to set-theoretic relations (Grygiel at al, 1999).

Normally, a structure requires the *complete* specification of its component relations. For example, AB:BC is defined information-theoretically by two distributions consisting of probabilities or frequencies for *all* (A_i,B_j) and (B_j,C_k), i.e., the full two tables shown in Table 1 (right). It is possible, alternatively, to define a model in terms of *any set of states* and their probabilities drawn from the top relation ABC and all of its projections (Jones, 1985). A small set of states might have probabilities unexpectedly high or low. These states represent salient "events" or "facts," and they, rather than complete projections, might be considered the "parts" of relation ABC. This approach is explained later in Section V. Analysis. This *state-based* approach is more powerful than the conventional *variable-based* framework, which it encompasses as a special case. The cost, however, of the state-based approach is a sizable increase in the size of the lattice of possible structures. As presented by Jones, the reference distribution for state-based modeling is the uniform distribution, but other reference distributions can also be chosen.

C. Complexity; Constraint

Structures differ in *complexity*, where this word can be given different meanings. In this article, the complexity of a relation or a distribution is the number of tuples or probability values. Decomposition reduces this complexity, i.e., is *compression*. This notion of complexity has the sense of randomness and is different from Wolfram's (1986) or Langton's (1992) "edge of chaos" complexity.

The complexity of a structure, information-theoretically, is its degrees of freedom which is the sum of the degrees of freedom of its relations, corrected for overlap (Krippendorff, 1986). For example, df(AB:BC) = df(AB) + df(BC) - df(B). (Defining a set-theoretic analog, however, is not straightforward.) df depends only on the variable cardinalities and not on the actual relations. For |A| = |B| = |C| = 2, df(ABC)=7 because a 2x2x2 table needs only 7 values to be specified (since probabilities sum to 1, and frequencies to the sample size). df(A:B:C) = 3 since only one probability value needs to be specified for each variable. df decreases by 1 at every level in Figure 3. Normalizing the df measure to a 0-1 scale gives

$$\text{Complexity}^{\text{norm}} = [\ df(AB:BC) - df(A:B:C)\ ] / [\ df(ABC) - df(A:B:C)\ ].$$

Complexity-reduction is achievable not only by decomposition of relations into simpler structures, i.e., by descending the Lattice of Structures, but also by the use of latent variables, which will be explained now, or by state-based modeling, which will be explained in the next section.

Complexity reduction can be accomplished by adding additional variables. A latent variable model typically *simplifies* the relation it explains. One cannot simplify in this way

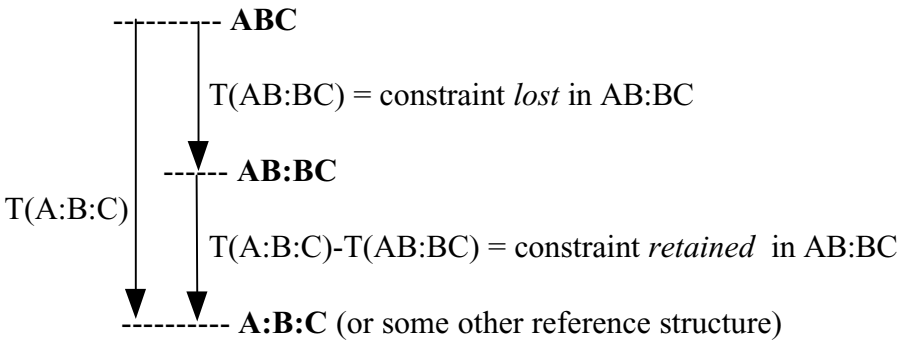
the relations of Table 1, so this approach will be illustrated with a different example. Consider an AC relation, latent variable B, and posited relation ABC having the structure BA:BC. If |A|=|C|= 4 and |B|=2, then df(AC) = 15 while df(BA:BC) = 13, so BA:BC is less complex than AC.

When considering some actual system, the lattice of structures indicates the possible decompositions of the top relation. Decomposition can be exact, i.e., with no constraint loss relative to ABC, or approximate, i.e., involving only small constraint loss. Where systems are highly decomposed, one can easily reduce the system to simple parts. There are systems, however, where the slightest decomposition results in *total* loss of constraint. This is illustrated by the set-theoretic relation ABC = {000, 011, 101, 110}, whose three dyadic projections, AB, AC, and BC are all heaps, i.e., {00, 01, 10, 11} which have no constraint at all. This can be visually represented by Figure 5.

Figure 5. Borromean rings (Removing one ring allows the other two to separate.)

Constraint loss increases monotonically as one descends the lattice (Figure 3). As indicated in Figure 6, the constraint *lost* in AB:BC is T(AB:BC) = U(AB:BC) - U(ABC) and the constraint *retained* is T(A:B:C) - T(AB:BC) (Figure 6) (Krippendorff, 1986). Constraint might alternatively be measured from a distribution other than A:B:C. For example, for directed systems where C is generated from A and B, the independence model is not A:B:C but AB:C.

Figure 6. Constraint lost and retained in structures.



Retained constraint is information *in* the structure. Normalizing to a 0-1 scale gives

$\text{Information}^{norm} = [T(A:B:C)-T(AB:BC)] / T(A:B:C) = 1-T(AB:BC)/T(A:B:C)$

The transmission measure allows a precise formulation of Simon's (1962) observation that most systems are "partially decomposable." Consider a system ABCD having two parts, AB and CD, and consider also the identity,

$T(A:B:C:D) = T(AB:CD) + [\ T(A:B) + T(C:D)\ ].$

Partial decomposability means that T(AB:CD), the *between*-parts constraint, is small compared to T(A:B) + T(C:D), the *within*-parts constraint. This is analogous to a between-within decomposition in the Analysis of Variance.

V. Analysis

A. Decomposition Losses

Structures encompass multiple relations, but every structure has an equivalent single (decomposable) relation linking all of the variables. This will be written as ABC_{model} while the data itself will be written simply as ABC. For example, $ABC_{AB:BC}$ is the calculated ABC distribution for model AB:BC. The equivalent relation for any model can then be compared to the data. Difference between the two is the error in the model (the constraint loss or information loss).

The single relation equivalent to a structure is essentially the same in both set- and information-theoretic formalisms: it is the relation with maximum uncertainty, given the constraints imposed by the model. For example, $ABC_{AB:BC}$ is the relation (distribution) which maximizes the uncertainty, U, subject to the constraints of AB and BC. These constraints appear on the right of Table 1 as the two lists of tuples or the two 2-variable frequency tables.

Set-theoretically, the equivalent relation is the intersection of all relations in the structure, each relation being first "expanded" by Cartesian products with variables missing in it; here $ABC_{AB:BC} = (AB \otimes C) \cap (BC \otimes A)$. For example, the equivalent triadic relation for the AB:BC model in the set-theoretic example of Table 1(a) is

$$\begin{aligned} ABC_{AB:BC} &= [\{00., \quad 01., \quad 11.\} \otimes \{..0,..1\}] \cap [\{ .00, \quad .10, \quad .11\} \otimes \{0..,1..\}] \\ &= \{000,001,\ 010,011,\ 110,111\} \quad \cap \quad \{000,100,\ 010,110,\ 011,111\} \\ &= \{000, \quad 010,011,\ 110,111\} \end{aligned}$$

Here, as noted above, $ABC_{AB:BC}$ is the same as the original ABC, so model AB:BC has no error or constraint loss. Neither AB nor BC can be dropped because this would yield an equivalent ABC relation having 6 tuples, while |ABC| = 5. (Note also that AC is a heap and adds no constraint.) The set-theoretic analysis for all models is summarized above in Table 3.

Information-theoretically, the equivalent single relation is the distribution $q_{AB:BC}(A,B,C)$ satisfying the conditions which maximizes

$U(AB:BC) = -\Sigma q_{AB:BC}(A,B,C) \log q_{AB:BC}(A,B,C)$

subject to the linear constraints: $q_{AB:BC}(A,B) = p(A,B)$ and $q_{AB:BC}(B,C) = p(B,C)$ Recall that p and q mean observed and calculated). Because model AB:BC is acyclic, the solution for this constrained maximization can be written directly; it is

$q_{AB:BC}(A,B,C) = p(A,B)\ p(B,C) / p(B)$.

For the information theoretic example of Table 1(b), Table 2 above gives the calculated distribution, $q_{AB:BC}(A,B,C)$ as well as the original observed probabilities, p(A,B,C).

For any structure the constraint of $ABC_{structure}$ is less than or equal to the constraint of ABC. Constraint loss is the "transmission" of the structure, also called cross-entropy or mutual information (there is an analogous set-theoretic expression). For the set-theoretic example above, the loss is 0, that is, the equivalent relation, $ABC_{AB:BC}$, is identical to ABC. For the information-theoretic example, the equivalent relation is not identical to the ABC, as can be seen by comparing the p and q distributions of Table 2. The loss is

$T(AB:BC) = -\Sigma\, p(A,B,C) \log [\ p(A,B,C)/q_{AB:BC}(A,B,C)\ ]$.

T can be computed directly from p (without first obtaining q) by T(AB:BC) = U(AB:BC) - U(ABC), where U(AB:BC) = U(AB)+U(BC)-U(B) and where the U's are computed from projections of the data. However, *cyclic* structures do not have algebraic expressions for U, and require the iterative generation of q and the calculation of T by the p log p/q expression. From T, normalized information is calculated. As shown above in Table 4, $Information^{norm}(AB:BC)= 0.98$. Very little constraint is lost in decomposing ABC to AB:BC. This, coupled with the greater simplicity of AB:BC compared to AB:BC:AC, is the basis for saying that AB:BC is the best model for the data of Table 1(a).

State-based information-theoretic structures are treated similarly. Calculated distributions, q(A,B,C), have maximum U, constrained not by entire projections of p(A,B,C) but by a set of selected individual probability values. The essence of how state-based decomposition can produce lower constraint loss is illustrated in Table 6.

Table 6. State-based decomposition

	B_0	B_1								
A_0	.1	.1	.2		.04	.16	.2		*.1*	*.1*
A_1	.1	.7	.8		.16	.64	.8		*.1*	**.7**
	.2	.8			.2	.8				
structure	AB			A:B				$[A_1,B_1]$		
df	3			2				1		
constraint loss, T	-			.087				0		

A:B, having probabilities p(A)*p(B), is not identical to AB and thus exhibits constraint loss. The AB model has df=3, and to show this 3 cells (arbitrarily chosen) are shaded. The A:B model has df=2, i.e., it needs only two specified probability values, one (arbitrarily chosen and shown shaded) from each margin. A state-based model specifying the *single* probability value, $p(A_1,B_1)$=**.7** (*not* arbitrary, shown shaded) forces the remaining (A_0,B_0), (A_0,B_1), and (A_1,B_0) probabilities, by the maximum uncertainty principle, to be *.1*. These are in fact correct and this state-based model thus has zero constraint loss even though it is simpler (has smaller df) than A:B. These data were of course "cooked" to produce this result, i.e., to show how a one-parameter state-based model could be superior to a two-parameter variable-based model. In the present example the state-based model has only one probability value, but in general such models can specify any (linearly independent) set of probabilities from an original table, its margins, its margins of margins, and so on. In the present example, [$p(A_0,B_0)$=.1, $p(B_1)$=.8] would be a legitimate df=2 state-based model.

State-based analysis (not shown here) of the earlier distribution of Table 1(b) reveals that a four-parameter state-based model, consisting of $p(A_1,B_0)$, $p(A_0,B_1)$, $p(B_0,C_1)$, and $p(B_1,C_0)$ would capture virtually the same amount of information as the five-parameter AB:BC model. (Note that the four states come from the AB and BC relations.) So, state-based analysis would improve upon the RA results summarized in Table 4.

B. Reconstruction and Identification

RA includes *reconstruction* and *identification*. The two examples of Table 1 illustrate reconstruction. In reconstruction, one starts from a whole, and decomposes it into provisional parts (relations) and then re-composes these parts to see if they account for the whole. In identification the parts are given, and one does only composition; this is done by the calculation of the equivalent single relation for the model, as discussed in the previous section.

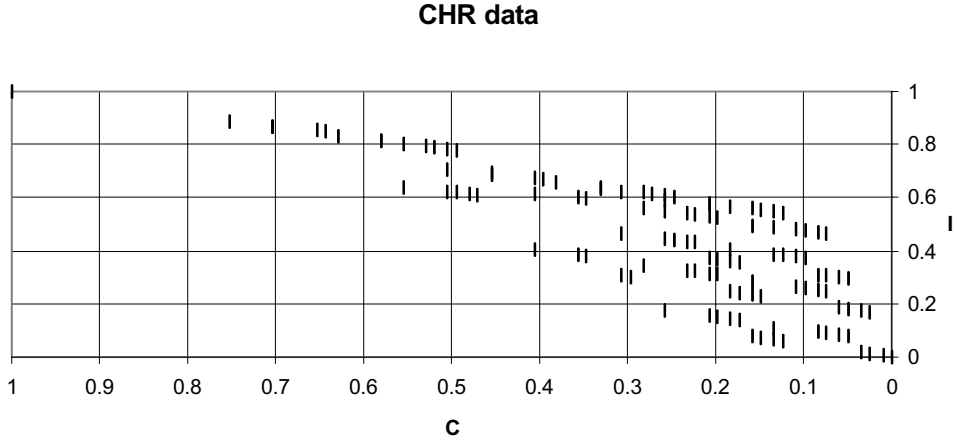
In reconstruction the objective is to find simple but low-error (high information) structures to model the data. There are different ways to balance the dual objectives of minimizing error and complexity. While conceptually one descends the Lattice of Structures and assesses the different decompositions, operationally one might actually either descend or ascend the lattice. If accuracy is the primary concern, or if systems are neutral, a modeling procedure might descend the lattice of structures until the error becomes too great. If simplicity is the primary concern, or if systems are directed, the procedure might ascend the lattice until further complexification is not forced or justified by the data. In information-theoretic/log linear modeling, which is Chi-square-based, the *statistical significance* of error, i.e., the probabilities of Type I and Type II errors, integrates error magnitude and complexity, but this does not resolve the issue: there are tradeoffs between these two kinds of errors (Knoke & Burke, 1980; Krippendorff, 1986).

Disallowing statistically significant error may prevent any simplification, so one might choose the simplest model whose error is acceptable. AB:AC:BC may be the simplest structure whose error is statistically *not* significant while the simpler AB:BC may have an

error which *is* statistically significant but only slightly bigger. Statistical significance is not pragmatic significance, and AB:BC might be preferred for its simplicity.

As already noted, picking a best *particular* structure does not constitute a complete analysis. A system is fully described by the constraint losses (or conversely, the information captured) for *all possible* decompositions. Table 4 above shows the complete reconstruction analysis of the data given in Table 1(b). Another illustration of a complete reconstruction analysis is provided by Figure 7 which plots $Information^{norm}$ against $Complexity^{norm}$ for data on medical and sociological characteristics of a sampled population (Zwick & Pope, 1999). The figure displays the entire lattice of possible decompositions for this data, treated as a neutral system. Models at the upper envelope of the cluster are plausible candidates for acceptance. That is, for any complexity, C, one wants a model with the maximum information, I.

Figure 7. Decomposition Loss Spectrum: ($Complexity^{norm}$, $Information^{norm}$). 114 four-variable structures; data from Kaiser Permanente Center for Health Research



A few words on identification. Structures can arise not from decomposition but from the composition of separate relations which may not be the projections of a single higher-ordinality relation. When wholes are thus composed of preexisting parts, there is no issue of constraint loss. If the relations overlap, e.g., AB and BC, they may be either consistent or inconsistent in their overlapping subrelation distributions, i.e., B. For example suppose one were given the AB distribution shown on the right side of Table 1(b), whose B margin is [1034, 444], but also an *altered* BC table whose B margin was, say, [1024, 454].. Such inconsistency could arise from sampling (or other) errors. When overlapping subrelations are identical, composition is straightforward, but if they are even slightly different -- and inconsistency is to be expected -- resolution of the inconsistency is first required (Klir, 1985; Anderson, 1996). Composition can also first utilize and then exclude extraneous variables in what might be called “reverse” latent class analysis. Suppose one is interested in the relation between A and C, but has inconsistent data only on AB and BC. After the inconsistency is resolved, the dyadic relations can be composed into a triadic ABC and then projected onto the desired AC relation (Anderson, 1996).

VI. Remarks

This article presents the essentials of Reconstructability Analysis, a particular set of procedures within general systems methodology (Klir, 1985). The framework presented here offers a very general approach to the multivariate modeling of nominal and quantitative data. This approach could be of significant use for analyzing the independencies among biological characters and genes and relevant attributes of the environment.

For further discussion on Reconstructability Analysis, see the special 1996 issue of the International Journal of General Systems on GSPS (General Systems Problem Solver). This framework is undergoing continued research and development, but tools which integrate what is already known are unfortunately not yet available. Ideally one wants a software implementation which can:

- do both set- and information-theoretic analyses and their fuzzy extensions
- using both variable- and state-based approaches
- in both confirmatory and exploratory (data mining) modes
- with efficient lattice search techniques for many variables
- on both nominal, ordinal, and quantitative data (and thus include effective binning)
- for static and dynamic (e.g., time series) applications
- with or without latent (supplementary) variables
- for both reconstruction and identification (including inconsistency resolution).

I am unaware of any software implementation which approximates these specifications, but the separate components exist. A software package aimed at this goal is being developed at PSU based on earlier PSU efforts and with external collaboration.

Acknowledgments

I thank George Klir for illuminating discussions on reconstructability analysis and systems theory, Anthony Blake for stimulating conversations on wholes and parts (and for the Borromean rings), and both them and George Lendaris, Bjorn Chambless, Jeff Fletcher, Stanislaw Grygiel, Michael Johnson, and Tad Shannon for helpful comments on the manuscript. I'm grateful to G∫nter Wagner for his patience and his gentle insistence that I try harder to make this article accessible. Given all this help, I am solely responsible for whatever obscurity still plagues this presentation.

References

Anderson, D. R. (1996). "The Identification problem of Reconstructability Analysis: A General Method for Estimation and Optimal Resolution of Local Inconsistency." Ph. D Dissertation, Portland State University. Portland, OR.

Ashby, W. R. (1964). Constraint Analysis of Many-Dimensional Relations. *General Systems Yearbook*, **9**, 99-105.

Bishop, Y. M., Feinberg, S. E., and Holland, P. W. (1978). "Discrete Multivariate Analysis." MIT Press, Cambridge.

Conant, R. C. (1981). Set-Theoretic Structure Modeling. *Int. J. General Systems*, **7**, 93-107.

Conant, R. C. (1988). Extended Dependency Analysis of Large Systems. *Int. J. General Systems*, **14**, 97-123.

Davis, J. A. (1985). "The Logic of Causal Order" (Quantitative Applications in the Social Sciences #55). Sage, Beverly Hills.

Grygiel, S., Zwick, M., and Perkowski, M. (1999). Multi-level Decomposition of Relations." (In preparation).

Hagenaars, J. A. (1993). "Loglinear Models With Latent Variables." (Quantitative Applications in the Social Sciences #94). Sage, Beverly Hills.

International Journal of General Systems (IJGS) Special Issue on GSPS (1996) **24** (1-2).

Jones, B. (1985). Determination of Unbiased Reconstructions. *Int. J. General Systems*, **10**, 169-176.

Klir, G. (1985). "The Architecture of Systems Problem Solving." Plenum Press, New York.

Klir, G. and Wierman, M. J. (1998). "Uncertainty-Based Information: Variables of Generalized Information Theory." Physica-Verlag, New York.

Knoke, D. and Burke, P. J. (1980). "Log-Linear Models." (Quantitative Applications in the Social Sciences Monograph # 20). Sage, Beverly Hills.

Knoke, D. and Kuklinski, J. H. (1982). "Network Analysis." (Quantitative Applications in the Social Sciences Monograph # 28). Sage, Beverly Hills.

Krippendorff, K. (1986). "Information Theory: Structural Models for Qualitative Data." (Quantitative Applications in the Social Sciences #62). Sage, Beverly Hills.

Langton, C. (1992). Life At the Edge of Chaos. In "Artificial Life II" (C. G. Langton, C. Taylor, J. D. Farmer, S. Rasmussen, Eds.), pp. 41-91. Addison Wesley, Reading.

Long, J. S. (1983). "Covariance Structure Models: An Introduction to LISREL." (Quantitative Applications in the Social Sciences #34). Sage, Beverly Hills.

McCutcheon, Allan L. (1987). "Latent Class Analysis." (Quantitative Applications in the Social Sciences #64). Sage, Beverly Hills.

Simon, H. A. (1962). The Architecture of Complexity. *Proc. Am. Philosophical Society*, **106**, 467-482. Reprinted in "The Sciences of the Artificial." (H. A. Simon)

Simon, H. A. (1981). "The Sciences of the Artificial." M.I.T. Press, Cambridge.

Wolfram, S (1986). "Theory and Application of Cellular Automata." World Scientific, Singapore.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control* **8** (3), 338-353.

Zwick, M. and Shu, H. (1996). Set-Theoretic Reconstructability of Elementary Cellular Automata. *Advances in Systems Science and Applications*. Special Issue **1**, 31-36.

Zwick, M. (1996). Control Uniqueness in Reconstructability Analysis. *Int. J. General Systems*, **24** (1-2), 151-162.

Zwick, M. and Shu, H. (1999). Reconstructability and Dynamics of Elementary Cellular Automata. Manuscript in preparation.

Zwick, M. and Pope, C. (1999). Reconstructability Analysis on Medical Utilization [OPUS] Data. Manuscript in preparation.