

Statistical entailments and the Galois lattice ¹

Douglas R. White

Institute for Mathematical Behavioural Sciences, University of California, Irvine, CA 92717, USA

Abstract

Statistical entailment analysis (White, 1984; White and McCann, 1988 *Social Structures: Form and Behaviour in Social Life*) (Cambridge University Press) pp. 380–404) aims first at a rigorous evaluation of null hypotheses of statistical independence as a potential source of binary data structure, and second at constructing a discrete structure (Boolean) model of those statistical interactions that remain when the null hypothesis is rejected for particular subsets of variables. Signal detection theory, rather than a conventional significance level, is used to specify optimal cutoffs given an ordering of ratios of actual to expected across levels of exception and relevance. Bivariate entailment analysis is generalized here to improve its utility for use in lattice approximation. Generalized statistical entailment analysis describes Boolean patterns in a set of data in terms of those that occur with greater frequency than expected by chance according to a model of complete statistical independence (the specific model of independence derives from a distribution of randomly permuted entries in the columns of the data matrix marginals, i.e. keeping univariate marginals fixed). This expands on the initial design of entailment analysis (White, 1984) to deal with partial orders of quasi-implication in pairs or chains of dichotomous variables, supported by statistical evidence of departure from bivariate independence and conformity to the rules of transitivity. Statistical approximations simplify a lattice representation of discrete structure by forcing quasi-implications (ignoring exceptions), for example, but they also provide information about those implications in the lattice that represent statistically significant tendencies. Given a lattice representing the discrete structure of a raw data matrix, the findings of entailment analysis describe additional structural regularities (tendencies towards further statistical constraints on Boolean patterns that occur in the data) that can be used to simplify (by approximation) the lattice of empirical patterns. As demonstrated with studies of dual orderings of material possessions (possessions stratify people; people possessions), the statistical interpretability of discrete structure lattices is enhanced by using the results of

¹The initial work in generalizing entailment analysis as described in this article was done under an appointment as Directeur d'Etudes Associé supported by the Maison de Science de l'Homme, Paris.

entailment analysis for consensus-simplification of statistically strong or significant implicational relations.

1. Introduction

Statistical entailment analysis (EA or SEA) is applicable where a propositional logic may describe patterns in binary data, and answers six classes of questions posed by problems of discrete structure. These are common problems in network as well as attribute analysis:

(1) What statistical interactions are significant? How best to evaluate significance not for single interactions, but for sets of interactions, or with small samples or low expected cell sizes?

(2) For each particular interaction among a subset of n variables (n -way interaction), what is the logical *form* of the relationship and its particular level of *exceptions*?

(3) What are the *symmetries* of the interactions? Are they best treated as asymmetric and discrete relationships such as $a \rightarrow b$ (a entails, b , but not the converse $b \rightarrow a$) or as symmetric and implicitly continuous (as in a covaries with b)?

(4) What is the statistical *relevance* of each interaction, as a comparison² of the effects of presence versus absence of the premise on conclusion outcomes?

(5) What is the nested *structure* of logical relationships in the data?

(6) How do we identify the structure of logical relationships given *exceptions or missing data*, which preclude the automatic formation of a logical structure?

Ordinary statistical analyses, because they assume a linear (for example, log-linear) model, do not treat questions 2 and 3. Most statistical analyses of n -way interactions such as log-linear analysis also require expected cell values > 5 , which in turn may depend on very large sample size. Questions of logical form also shape the answers to questions 5 and 6 in a way not ordinarily treated as a statistical problem.

This article presents a *generalized entailment analysis* (EA+) that treats n -way interactions as a problem in discrete structure. The generalization develops out of entailment analysis of bivariate relationships or first order implications (EA1),³ a measurement model for multi-dimensional Guttman scaling (White, 1984). Statisti-

² Since every implication a entails b has an equivalent contrapositive form not b entails not a , with reversal of premises and conclusions, the comparative measure of relevance is necessarily in symmetric form, i.e. as a correlation or relative product coefficient.

³ The EA1 model assumes no higher order intersections, but null hypothesis tests of three-way or higher order interactions among the variables (White et al., 1983) and appropriate group tests of interactions (White and Pesner, 1983) are used to ascertain the appropriateness of this assumption. In the generalization that follows, models of higher-order interaction use these same independence models to test for rejection of the null hypothesis.

cal criteria for EA + are the same as given in White and McCann (1988, pp. 382–384; see White et al., 1977 for a sample application) for EA1, as explained below.⁴ After providing more detail on the statistical criteria used in entailment analysis, an analysis of statistical entailments in data on material possessions will exemplify how entailment and lattice analysis may complement one another. The Galois lattice is an appropriate means of representing generalized entailment structures.

2. Generalized entailment analysis (EA +)

N-ary statistical relationships or *n*-way interactions are examined in entailment analysis as *quasi-implications*: raw empirical relationships treated in the form of Boolean logic, but *quasi* in that they require evaluation of *exceptions*, *statistical relevance*, and *significance* (including questions of generalizability given sampling or measurement variation) under group tests of the null hypothesis. For a set *X* of 0/1 variables with or without missing (coded ‘.’) values, let the Boolean intersection of a subset of variables, e.g. {*a, b, c*...} be *x* with values $x = \{ \cdot \}$ where any missing data are present; 1 iff $a = 1, b = 1, \text{ and } c = 1 \dots$; and otherwise 0. For set *z* intersection values, the *complement* $z' = \{ 1 \text{ iff } z = 0; 0 \text{ iff } z = 1; \cdot \text{ iff } z = \cdot \}$. A quasi-implication such as $x \rightarrow y$ (in raw form: $a \wedge b \wedge c \dots \rightarrow y$, read: the conjunction or *intersection* of *a, b, c*... imply *y*) may have *exceptions* xy' , where $x = 1$ ($a \wedge b \wedge c \dots = 1$) but $y = 0$. *Cardinalities* of sets defined by variables such as *x, x* \wedge *y, x* \wedge *z'* are denoted $|x|, |x \wedge y|, |x \wedge z'|$. Thus $x \rightarrow y$ has no exceptions if $xy' = |x \wedge y'| = 0$.

Every quasi-implication can be evaluated in its own right for exceptions and statistical relevance – question 2 and 4 above. Question 1, 3, 5, and 6, however, require that we evaluate each quasi-implication with respect to the entire system of variables that are in play in a particular analysis. As answers to these questions, we establish a generalized set of criteria for entailments (White and McCann, 1988):

A *exceptions* (Q2). $x \rightarrow y$ but not ($y \rightarrow x$) implies there are fewer exceptions xy' ($x = 1, y = 0$) than exceptions yx' ($y = 1, x = 0$): hence $|x \wedge y'| < |x' \wedge y|$.

B *relevance* (Q4). $x \rightarrow y$ implies that $x = 1$ increases the probability of $y = 1$ and contrapositively $y = 0$ increases the probability of $x = 0$: hence $r_{xy} > 0$ (positive correlation).

C *transitivity* (Q5, 6). $x \rightarrow y$ and $y \rightarrow z$, for logical consistency, require that *y* does not diminish the conditional probability of *z* given *x*: hence non-negative partial correlation, where are no *x, y, z* such that $x \rightarrow y \rightarrow z$ and $r_{xz,y} = [r_{xz} - r_{xy}r_{zy}] / ([1 - r_{xy}^2][1 - r_{zy}^2])^{1/2} < 0$.

D *non-independence* (Q1, 3). The aggregate frequency of observed entailments at a given level of exceptions, relevance and form is greater than the expected

⁴ Programs for entailment analysis are available from the author at the School of Social Science, University of California, Irvine, CA 92717, USA.

frequency under the assumption of total independence of the variables.

E replication with respect to control variables for a given form of an implication [Q2] establishes the lack of higher order interactions: hence if $x \rightarrow y$ then for all z and their complements z' , $|z \wedge x| > 0$ implies $z \wedge x \rightarrow y$ and $|z' \wedge x| > 0$ implies $z' \wedge x \rightarrow y$.

Unary entailment analysis (EA1) with single premises (White, 1984; see White and McCann, 1988) computes a distribution of 2×2 tables (excluding missing data) classified by strength of correlation and percentage of exceptions to quasi-implication. For example, a quasi implication $x \rightarrow y$ might have 2% error xy' in a sample size N and a relevance coefficient $r_{xy} > .6$. Generalized (n -ary) entailment analysis (EA+) includes the statistical evaluation of quasi-implications with unary and multiple premises.⁵ Just as the statistical distributions of entailments in unary premises are generated by evaluating all 2×2 tables holding their univariate marginals constant, so binary premises are evaluated in $2 \times 2 \times 2$ tables holding the 2×2 marginals constant, trinary premises are evaluated in $2 \times 2 \times 2 \times 2$ tables holding the $2 \times 2 \times 2$ marginals constant, and so forth.

To evaluate non-independence (Q1), frequency distributions of tables classified by relevance and exception levels are compared to expected frequency distributions. The default method is multiple (Monte Carlo) randomization of the raw data with fixed raw marginals to derive sample estimates for the expected distribution.⁶ The variables of the data matrix are randomized and analyzed for quasi-implications a sufficient number of times to give a reliable distribution of expected frequencies for *four logical types* of quasi-implication: strong and weak inclusion, coexclusion and coexhaustion. For every pair of variables x and y (where x is expanded in EA+ to include set intersection values),⁷ actual or randomized, there are two quasi-implication candidates. These are *strong and weak inclusions* (if x then y ; if y then x , *strong* being that with fewer exceptions) if the correlation r_{xy} is positive, but if the correlation r_{xy} is negative one candidate is *coexclusion* (if x then not $y \equiv$ if y then not $x \equiv x \rightarrow y' \equiv y \rightarrow x'$) and the other *coexhaustion* (if not x then $y \equiv$ if not y then $x \equiv x' \rightarrow y \equiv y' \rightarrow x$). Expected frequencies of values for each type are computed for different rates of exception

⁵ The entailment programs that do higher order interaction tests (criterion 5) are currently separate from the EA1 program that does the bivariate analysis of entailments with many premises. Use of the procedures for evaluating interactions (n -ary implications) in lattices may help to combine the n th criterion into a single constructive procedure that generates successive layers of n -ary entailments ordered by levels of exception.

⁶ Each variable string of $\{0, 1, '\cdot'\}$ values is considered as a sampling distribution for random reassignment of entries. Sampling without replacement generates a random reordering of values. With or without replacement, a set of randomly simulated variables has the same expected distribution of marginal frequencies as the actual data. An exact method of determining expected frequencies in n -way interactions, useful for small samples, uses the hypergeometric distribution (White et al., 1983) to compute an expectation based on each possible set of cell values given marginal totals.

⁷ In current implementations the analysis of 3-way and 4-way interactions, holding marginals constant, are done separately from EA1. In principle EA1 could be expanded directly into an EA+ algorithm by generating premise sets as intersections of 2 or more actual or random variables.

and different levels of association between the variables, and compared with the actual distribution of quasi-implication by type, level of exception and level of association r_{xy} . Note that since EA uses randomization methods, it is neither dependent on distributional assumptions nor sufficient sample size for test validity.

Entailment analysis retains that set of quasi-implications (possible logical propositions, with possible exceptions) that optimally describes the non-independent components of a data structure. The method is constructive. Within each logical type, quasi-implications are sorted from 'stronger' to 'weaker' by discrete levels of exceptions and, within each level, by classes of relevance (Q4). What are the optimal cutoffs along these rankings to determine a level above which the relation between premises and conclusions of 'stronger' quasi-implications are considered non-independent, and below which 'weaker' ones are rejected under the null hypothesis of independence between premises and conclusions? The cutoff problem is one where the theory of signal detection applies.

Signal detection theory (Coombs et al., 1970) treats the statistical decision problem of optimal cutoffs as to where actual frequencies differ from expected, such as those of quasi-implications ordered by relevance at a given level of exceptions, given their relative distributions. At each level of exceptions, *which relevance level is the cutoff* (constrained to be non-decreasing in relevance as we move to weaker exceptions) *above which* the non-independence hypothesis is optimally probable as opposed to that of statistical independence? Once such cutoffs are determined, each level of exceptions will have its own actual/expected ratios, and signal detection is used again to determine a cutoff exceptions level, *below which* the likelihood of non-independence is optimal.

The question of *optimal* statistical decisions in signal detection is different than which probability level to use in significance tests. If the 'null' or 'background noise' model is analogous to distribution characteristics of friendly aircraft against which a military does not want to fire, and our decision makers in the analogy have an observed set (distribution) of incoming aircraft signals, the problem is not whether to fire at any given aircraft judged as hostile ('signal') at some preassigned significance level of wrong assessment, such as $p < 0.01$. This is only type I one of two types of statistical decision error, and in this case may let in too many hostile aircraft, the type II error of treating 'signal' as noise. Weighting of acceptable risk probabilities of wrong decisions of types I (false positives) and II (undetected signals) is balanced by specific alpha and beta 'risk aversion' parameters.

Actual/expected cutoff ratios greater than 1.0 in our signal detection problem help to insure that quasi-implications are not due to random instability in measurement (unreliability) or sampling (small sample size, standard errors under resampling). Increasing the cutoff ratio decreases the risk of accepting false positives, but increases the risk of undetected signals, that of rejecting valid quasi-implications. Entailment analysis describes an actual to expected ratio distribution by smoothed fit to the parameters of a gamma function which may peak on the side of high actual to expected ratios, and trail off on the side where expected exceeds actual. The optimal point to separate signal from noise in balanced statistical decisions (equal weighting to types I and II errors) is after the excess peak of actual/ex-

Table 1
Coefficient for optimizing entailments of a given logical type

Strong Inclusion (at a given exception level)	Expected	Actual
Candidates included in set	<i>a</i>	<i>b</i>
Candidates excluded in set	<i>c</i>	<i>d</i>

pected, where the derivative of the curve changes sign and the ratio of noise increases relative to that of undetected signals. Automated determination of optimal ratios for signal/noise judgments is a default option in the statistical entailment algorithm.

Other methods to settle the signal detection problem include assigning relative weights for the seriousness of type I versus type II errors, or maximizing an objective fitness function that varies by exception level, applied to all possible candidates for each of the four types of quasi-implication. One such objective function is a correlation applied to the relative frequency of expected versus actual occurrence of pairs of variables that are included or excluded in the set of propositions. Cutoffs based on the level of exceptions which maximizes the Pearson Product-moment correlation based on four cells (*a*, *b*, *c*, *d*) for strong inclusions, as in Table 1, are repeated for weak inclusion, coexclusion, and coexhaustion. This method minimizes risk of decision errors while maximizing the objective fitness function for actual/expected values in final choice of elements for the set of structure-describing propositions in a set of data. A binomial test for *k* observed implications in *n* trials with expected probability *p* may also be employed. To determine optimal levels at which to accept or reject the null hypothesis, however, signal detection methods, applied to actual/expected frequencies are preferable to arbitrarily chosen significance tests.

With respect to an overall structure of logically consistent entailments (Q5, 6), the constructive method of EA takes the best of the quasi-implication candidates (best meaning first, those with fewest exceptions, and second, those with the strongest correlation), then iteratively tests the next best candidate for transitivity with all previously accepted candidates, rejecting the candidate if it is intransitive ($x \rightarrow y \rightarrow z$ is intransitive if the partial correlation $r_{xz,y}$ is negative) or if the actual frequency of this class of quasi-implication is judged by signal detection as consistent with the null hypothesis.

Since EA adds implications in order of rank (exceptions level first) to an existing set of propositions, provided they pass the transitivity test, the resultant entailment structure contains two partial orderings. The first, at any given level of exceptions, is a partial ordering of inclusion relations among Boolean variables. The second ordering is that between any two sets of propositions that differ as to whether they are true for two different levels of exceptions. Smaller sets of stronger entailments are contained in larger sets of weaker ones. For a given problem, one can use the default (statistically optimal) cutoffs for exceptions, or select a substantively appropriate cutoff for the level of exceptions. In any case,

there may be different layers within the entailment structure (each logically consistent) according to the level of exceptions.

Whether the resulting entailments are symmetric or asymmetric (Q3) depends on the outcome of analysis, which is not weighted in one direction or the other. Unlike many of the discrete structure methods of analysis, entailment analysis is not dependent on judgments as to implicational asymmetry based on a priori decisions or arbitrary cutoffs as to tolerated levels of exceptions.

Results of an EA1 entailment analysis are reduced to their most parsimonious logical form by converting the three logical types of implication (\rightarrow inclusion, \rightarrow' exclusion, $'\rightarrow$ coexhaustion) into two types (inclusion, exclusion), with limited complementation of certain elements. Rather than such subsets of elements where $x \rightarrow y$ (inclusion), $y' \rightarrow z$ (coexhaustion), and $x \rightarrow' z$ (exclusion), certain elements are complemented to eliminate coexhaustions. In this case the equivalent structure is $x \rightarrow y \rightarrow z'$, where z is the complemented element. Thus, any given coexhaustive such as $x(' \rightarrow)y$ might equate to $x' \rightarrow y$. There are a minimum number of such complementations required of any first order predicate calculus to eliminate coexhaustives.

3. Galois lattices as representations of entailment structure

To represent discrete logical (e.g. Boolean) relations in binary data such as a raw 0/1 data matrix, the Galois lattice is a nearly perfect mathematical structure. It is ideal to represent structures of discrete logic (Guiges and Duquenne, 1986; Duquenne, 1987) in terms of the dual ordering of row and column relations (intersection, set inclusion and row-profile to column-profile correspondences) in binary rectangular data matrices (Barbut and Monjardet, 1970). The Galois lattice of sets of column elements formed by their closure under intersection is inversely isomorphic to the partial ordering of row elements by closure under intersection (Duquenne, 1992). The complementation and union operators, however, are missing from the Galois lattice.

The Galois lattice can represent the results of entailment analyses in various ways, either concretely or abstractly oriented, as the following:

(1) Concrete, error-free, two-mode (row *and* column) entailment analyses of a 0/1 matrix are expressed as implications among subsets of row elements, reading inclusion relations 'up' the Galois lattice, and implications among subsets of column elements as inclusions 'down' the lattice. Here, EA + merely expresses which implications are of statistical interest but does not alter the structure of the lattice.

(2) Abstractly, one-mode first-order entailments can be expressed in Galois form taking as input the matrix of implications among column variables. This is the simplest lattice (the Dedekind–MacNeil simplification) that contains the implicational partial ordering. There is no way, however, to express two-mode entailments in this fashion.

(3) The concrete two-mode lattice, as Duquenne shows in the article following this, can be saturated with additional implications (inclusion relations between sets of row elements or their intersections, likewise for column elements) resulting from statistical data analysis, simplifying certain aspects of the lattice by eliminating statistically less-common-than-expected Boolean patterns and ignoring exceptions to quasi-implications. Duquenne's syntax for lattice approximation by 'consensus simplification' re-situates a lattice element representing the premise of a quasi- or approximate entailment (as a meet of intersected variables) so that is ordered beneath (i.e. implying) the meet of its approximate conclusions, and implied reorderings are adjusted accordingly. (An alternate 'independence simplification' approach, yet to be worked out in terms of lattice algebra, is to *add to the lattice* of empirical possibilities those rare Boolean patterns which may fail to occur in small samples but that may be derived from statistical evidence of independence.)

The Galois lattice, however, unlike its more complicated companion the Boolean lattice, does not contain complement or union operators and thus cannot express coexhaustion relations among elements, where it is x or y that may exhaust the universe (or another set). It is only because EA1 *eliminates* coexhaustion relations, replacing certain elements with their complements, that nothing is lost in the Boolean logic of relations. Inclusion and exclusion relations can be read directly from the Galois lattice, the latter by elements that intersect at the zero cardinality (empty) element where $x \rightarrow y'$ is expressed as $|x \wedge y| = 0$.

In the case of a generalized entailment analysis (i.e. of n -way Boolean interactions), however, it is not the case that complementation of elements can always achieve elimination of coexhaustives. A certain care must be exercised, in interpreting the full array of n -way interactions, not to reduce the focus of the analysis simply to set inclusions and exclusions if significant coexhaustions are indicated from generalized entailment analysis (EA +).

4. What entailment can do for Galois lattices and vice versa

What entailment analysis can do for Galois lattice representations of discrete structure in empirical data is to offer an additional analytic capability and meaningful simplifications or to raise questions as to interpretation and generalization. Eliminating coexhaustions and replacing them by logically equivalent Boolean complements of certain variables in a discrete structure analysis can be a valuable step in simplification of a data structure. This may be achieved by entailment analysis prior to the use of Galois lattice representation of findings. For example, in a questionnaire in which there are a number of cumulative Guttman scales among items, but many of the questions that form implicational chains are negatively worded, Galois lattice expression of the raw data would break each chain into two, treating the positively and negatively worded elements as separate implicational chains. Thus, $a \rightarrow b' \rightarrow c \rightarrow d' \rightarrow e \rightarrow f'$ would become $a \rightarrow c \rightarrow e$ and

$f \rightarrow d \rightarrow b$, with exclusion between a/b , c/d and e/f . The simplicity of the structure is lost in the raw lattice, but not if elements b , d and f are complemented first, and the transformed data are represented in Galois form. EA1 provides a unique solution to the problem of the minimum number of elements to complement to capture, in Galois form, a discrete data structure with limited Boolean elements involving complementation or union.

Another reason for using entailment analysis with empirical data prior to lattice analysis is that the raw lattice may contain such a high proportion of the possible Boolean intersections that some kind of simplification is necessary. Duquenne's 'consensus simplification' corresponds to what is likely to be recurrent in successive replications of data collection with similar sample size. It is a descriptive and summary approach to discrete structure, retaining particularities in the data while adding statistically valid quasi-implications. Consensus simplification will be illustrated in the examples below.

To show implications of statistical significance (strong or significant n -way interactions as inclusions or quasi-implications) in a consensus-simplified lattice, covering relations could, for example, be colored or darkened in contrast to non-significant orderings. Remaining coverings in the lattice would then express the intersections of statistically independent variables.

When missing data are present in a set of 0/1 data subjected to discrete structure analysis, a number of problems of representation and interpretation arise. Entailment analysis provides a relatively simple solution: the statistical analysis is done on the data that are available, and findings are checked for the logic of transitivity to make sure that logical inconsistencies do not arise. Where they might arise, in the constructive approach, the weaker of two inconsistent statistical 'signals' is rejected, and the stronger retained.

Thus, entailment analysis deals with the problems of missing data, 'noisy' structure, and approximations (exceptions) as well as exact discrete structure, and finds significant simplifying implications for a modifying a Galois lattice representation in one of two ways. The first, Duquenne's 'consensus simplification', saturates the lattice with additional implications by re-situating elements to express new conclusions. The second is the entailment (EA1) option of filling-in missing data and/or replacing exceptional values in the data matrix from statistical entailments, using the revised data structure as input to Galois lattice analysis.

What lattices can do for entailment analysis, besides providing a means of representation, is to provide the machinery to check transitivity in the generalization of statistical entailment to include n -ary implications (and, for example, improve the efficiency of the entailment algorithm in not having to repeat an n -way interaction test if limiting marginal constraints have already been discovered at level $n - 1$). In adding new entailments to the basis by re-situating elements, a revised strategy for making entailment analysis more efficient in relation to discrete structure analysis is to call a saturation algorithm inside entailment analysis that takes a new candidate for n -ary implication and then checks its implications (Duquenne, 1991) for statistical transitivity. If consistent, the candidate is accepted, and the basis of implication is updated.

5. Illustrations: Tahitian urbanities' and Mbuti hunters' material possession lattices

Material possessions that stratify their possessors into a social structure, and reciprocally, people who order their possessions into 'hierarchies of acquisition', constitute a dual ordering of which the Galois lattice is a useful representation. In the domain of material possession studies, Schweizer (1993b; pp. 472–478) presents several illustrations of Galois representation of discrete data structure.⁸ Galois lattices of possessions among Papeete Tahitians and Mbuti hunters in Figs. 1 and 2, where people are stratified by their possessions, and possessions stratified by their possessors, are informed and simplified by entailment analysis results that modify Schweizer's original Figs. 3 and 4. At the top of the lattices in Figs. 1–4 are the common necessities (Papeete bicycles and primus stove; Mbuti machetes and hunting nets) held by people lower down in the lattice; lower down are the more privileged persons cumulating rare possessions in addition to common ones. The examples are sufficiently simple so that a reader can assess the credibility of the entailment simplification.

Statistical entailment for Papeete finds no significant three-way interactions, but additional unary premise implications (not in Fig. 3) are significant with single exceptions:

$F \rightarrow K \rightarrow R$ (refrigerator \rightarrow kerosene stove \rightarrow radio)

$F \rightarrow V \rightarrow R$ (refrigerator \rightarrow scooter or motorbike \rightarrow radio)

$B \leftrightarrow P$ (bicycle \leftrightarrow primus stove; each direction of implication has an exception).

For the Mbuti, entailment analysis gives a single significant three-way interaction, one that is shown in the lattice of Fig. 4 as well: only in the presence of dogs (used for hunting) do bows and hoes (also used for hunting) tend to co-occur (the meet or downward intersection or binary implication of bows B and hoes h implies d , dog). A number of additional unary premise implications (not in Fig. 4) are statistically significant with single exceptions (EA1):

$h \leftrightarrow d$ (hoe \leftrightarrow dog, both used in hunting; each direction of implication has an exception).

$C \rightarrow A \leftrightarrow o \leftrightarrow k$ (utensils: metal Cup \rightarrow anvil \leftrightarrow small mortar \leftrightarrow tusk hammer).

$s \rightarrow I$ (shovel \rightarrow iron pan for gold mining).

When the lattices in Figs. 3 and 4 are saturated by Duquenne's method (this issue) with these additional implications, we get the 'consensus simplification' in Figs. 1 and 2. As compared to Fig. 3, simplifications of the lattice in Fig. 1 result from re-situating F to its intersection under V (since $F \rightarrow V$ with one exception), then V under R (since $V \rightarrow R$ with one exception) so that we now have the ordering $F \rightarrow V \rightarrow R$. However, since $F \rightarrow K$ with one exception we move F further down under K and since $K \rightarrow R$ with one exception another ordering is generated where $F \rightarrow K \rightarrow R$. Finally, since $B \rightarrow P$ are reciprocally $P \rightarrow B$, each with one exception,

⁸ Interesting properties of the lattice structure of Schweizer's (1993a,b) 'cultural consumption' lattice for a peasant village in rural Java were first discovered using the entailment approach.

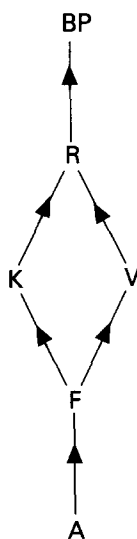


Fig. 1. Lattice of the dual ordering of consumer durables and households in Papeete, simplified by entailment analysis. A automobile; B, bicycle; F, refrigerator; K, kerosene or gas stove; P, primus stove; R, radio; V, two-wheeled motor vehicle.

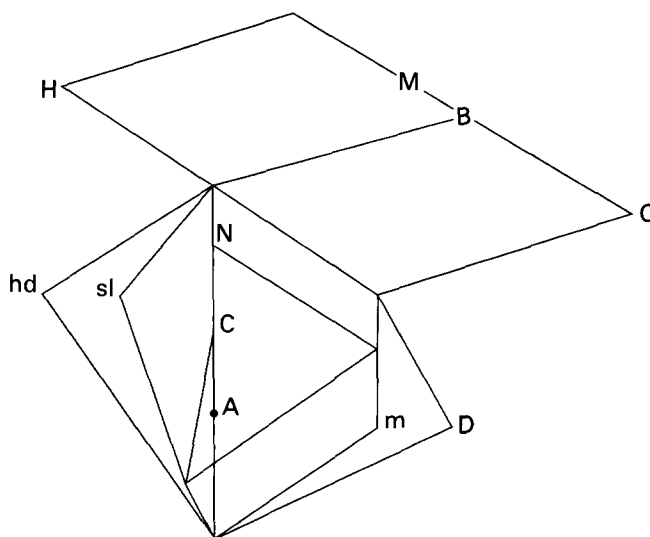


Fig. 2. Lattice of the dual ordering of material possessions among Mbuti hunter gatherers, simplified by entailment analysis. A, anvil; B, bow; C, metal cup; D, drum; d, dog; H, hunting net; h, hoe; I, iron pan for gold mining; k, tusk hammer; M, machete; m, bamboo flutes; N, spoon; O, large mortar; o, small mortar; s, shovel.

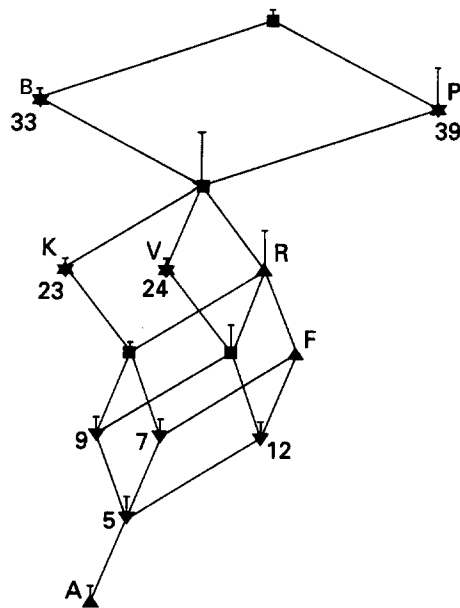


Fig. 3. Lattice of the dual ordering of consumer durables and households in Papeete, unsimplified. A, automobile; B, bicycle; F, refrigerator; K, kerosene or gas stove; P, primus stove; R, radio; V, two-wheeled motor vehicle (source: T. Schweizer 1993, p. 474).

B and P move downward to meet in their new position together at their former intersection. Hence the resultant logical simplification of quasi-implications into Fig. 1. K and V are not ordered with respect to one another, however, since there are four cases of K without V, and seven of V without K (the raw data table may be found in Schweizer, 1993; p. 471).

The entailment-simplified logic of material possessions in the examples shown in Figs. 1 and 2 is fairly transparent. For Papeete, a single status dimension of acquisition ordering emerges. For the Mbuti, the structure is still complex, but coherent subdimensions of the entailment-simplified lattice are interpretable. The ordering $O \rightarrow B \rightarrow M$ on the right are the subsistence-tool possessions (O = large mortar implies B ow implies M achete); H on the left is the hunting net. Below the H – B intersection are elaborations of further hunting technology ($h \leftrightarrow d$), mining ($s \leftrightarrow I$), and tools/utensils ($A \rightarrow C \rightarrow N$). Below the HB – O intersection are two independent musical possessions (D, m), the latter implying the lower ordered utensil, N . Six differentiated positions of persons, no one covering any of the others, are elaborated at the bottom of the lattice, with fewer more impoverished positions at the top held by person who possesses only certain of the widely shared elements. While simplified, this structure still shows “fundamental and fine distinctions of the division of labor in hunter–gatherer community: first a basic contrast between hunting techniques, then considerable sharing of types of special possessions, and last some wealth differentiation and occupational specialization in

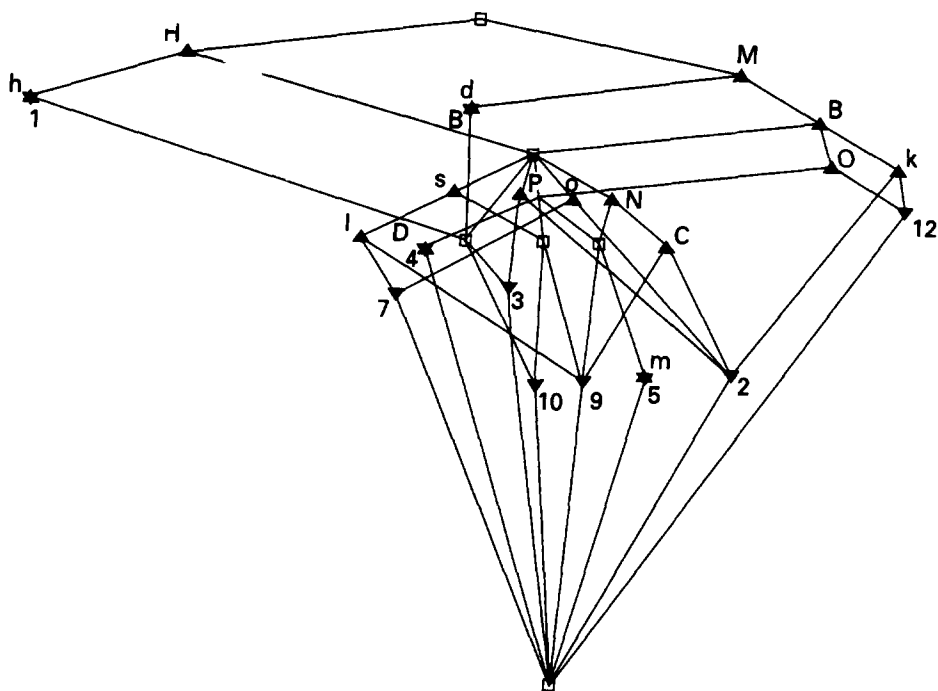


Fig. 4. Lattice of the dual ordering of material possessions among Mbuti hunter-gatherers, unsimplified. A, anvil; B, bow; C, metal cup; D, drum; d, dog; H, hunting net; h, hoe; I, iron pan for gold mining; k, tusk hammer; M, machete; m, bamboo flutes; N, spoon; O, large mortar; o, small mortar; s, shovel (source: T. Schweizer 1993, p. 477).

the set of rare possessions” (Schweizer, 1993b; p. 476). The Mbuti show a more egalitarian bias toward a differentiated structure (with one two-premise implication) on a shared base,⁹ quite different in form than the more unidimensionally property-stratified Tahitian township.

6. Summary and conclusion

Taking into account the possibility of random measurement error, missing values, and exceptions, statistical entailment analysis extracts structures of discrete logic from empirical data. While the details of the method are given elsewhere (White and McCann, 1988), the generalization of entailment analysis developed here is an ideal tool to complement Galois lattices in the representation of discrete

⁹ While I have not presented any of the actual entailment tables comparing actual to expected frequencies of the implicational relationships in these structure, the Mbuti data are much closer overall to a global null hypothesis than the more highly structured Papeete data.

structure in empirical data. When missing data are present, or the raw data contain significant measurement error, or there are exceptions to significant empirical implications, the Galois lattice representation of exact discrete structure may prove too complicated to use productively. Boolean analysis offers another statistical solution that begins from a more theory-driven concern with measurement, tested statistically in terms of models of discrete structure (Degenne and Lebeaux, this issue). Boolean and entailment analysis are complementary statistical approaches to hypotheses about discrete structure. The first is based on model-driven criteria for non-independence, the second on optimal signal/noise detection in accepting or rejecting null hypotheses about overall structure as well as components of discrete structure. Both identify statistical interactions in terms of Boolean patterns of variables that are disallowed or unlikely and can be described in terms of entailments or quasi-implications with exceptions. For example, where cases may be present with variables a and b but not without c , the 'disallowed' Boolean pattern corresponds to the quasi-implication $ab \rightarrow c$. Disallowed Boolean patterns are thus in one-to-one correspondence with sets of quasi-implications (n -ary premises their conclusions), and this correspondence extends to the Boolean or Galois lattices constructed from the data matrix. The Galois lattice¹⁰ is constructed by inclusion ordering of the closed sets of variables whose column profiles are formed by intersection, which corresponds dually to orderings of row profile intersections.

Duquenne (this issue) develops a syntax for Galois lattice approximation, and shows how to revise lattice representations of discrete structure in empirical data using the statistical findings of entailment analysis. Consensus simplification of discrete structure lattices provides a theory of re-situating elements in the lattice on the basis of statistical evidence. This may improve our interpretation of both the descriptively particular elements in a set of data as well as the statistically important quasi-implications. In the domain of material possession studies, for example, empirical examples of lattice simplifications that add implications from statistical entailment show improved clarity and interpretability as discrete structure approximations.

Consensus simplification is not the only theoretical basis for the lattice approximation problem, but it does correspond to what is likely to be recurrent in successive replications of an empirical study with comparable observations and similar sample size. A second theory is needed for Boolean expansions of rare event combinations on the assumption that findings were to be extrapolated to finite observations from larger sampling frames. For both types of lattice approximations, statistical entailment analysis can provide results that help us understand the statistically salient features of discrete data structures.

¹⁰ The Boolean lattice, of course, adds complementation and union operators to the algebra generating the sets of column elements, but thereby loses the ability possessed by the Galois lattice to represent implications dually: both as orderings among column element intersections and row element intersections. Thus, the Galois lattice, which is also much simpler than the expanded Boolean lattice for any given set of data, is generally preferable for interpreting the logic of discrete structure.

References

- Barbut, M. and B. Monjardet, 1970, *Ordre et classification*, 2 vols., (Hachette, Paris).
- Coombs, C.H., R.M. Dawes and A. Tversky, 1970, *Mathematical psychology* (Prentice-Hall, Englewood Cliffs).
- Duquenne, V., 1987, Contextual implications between attributes and some representation properties for finite lattices, in: W. Ganter ed., *Beitrage zur Begriffsanalyse* (Wolfe, Mannheim) 213–239.
- Duquenne, V., 1991, On the core of finite lattices, *Discrete Mathematics* 88, 133–147.
- Duquenne, V., 1996, On lattice approximations: Syntactic aspects, *Social Networks* 18, 189–199.
- Freeman, L.C. and D.R. White, 1993, Using Galois lattices to represent network data, in: P.V. Marsden, ed., *Sociological methodology* (Blackwell Press, Cambridge, MA) 127–146.
- Guiges, J.-L. and V. Duquenne, 1986, Families minimales d'implications informative résultant d'un tableau de données binaires, *Mathématiques et Sciences Humaines* 95, 5–18.
- Schweizer, T., 1993a, The cultural use of things: Consumption in rural Java, *Indonesia Circle* 61, 3–20.
- Schweizer, T., 1993b, The dual ordering of actors and possessions, *Current Anthropology* 34, 469–483.
- White, D.R., 1984, *Statistical entailment analysis*. Microcomputer Version for the IBM PC (PDP-10 Version 1975) (School of Social Science, University of California, Irvine, CA.)
- White, D.R. and H. Gilman McCann, 1988, Cites and fights: Material entailment analysis of the eighteenth-century chemical revolution, in: S. Berkowitz and B. Wellman, eds., *Social structures: form and behavior in social life* (Cambridge University Press, London) 380–404.
- White, D.R., M. Burton and L. Brudner, 1977, Entailment theory and method: A cross-cultural analysis of the sexual division of labor, *Behavioral Science Research* 12, 1–24.
- White, D.R. and R. Pesner, 1983, Internal replication and the systems concept in experimental research, *Behavioral Science Research* 18, 26–44.
- White, D.R., R. Pesner and K. Reitz, 1983, An exact significance test for three way interactions, *Behavioural Science Research* 18, 103–122.